

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2003, 04:14:55 ; Search time 111 Seconds  
(without alignments)  
5765.816 Million cell updates/sec

Title: US-09-701-229-1  
Perfect score: 1450  
Sequence: 1 cgtgctgctgctgctgcca.....tcttgagcggcagcgcatc 1450

Scoring table: OLIGO.NUC  
Gapop 60.0 , Gapext 60.0

Searched: 569978 seqs, 220691566 residues

Word size : 0  
Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued\_Patents\_NA.\*  
1: /cgn2.6/ptodata/1/ina/5A\_COMB.seq.\*  
2: /cgn2.6/ptodata/1/ina/5B\_COMB.seq.\*  
3: /cgn2.6/ptodata/1/ina/6A\_COMB.seq.\*  
4: /cgn2.6/ptodata/1/ina/6B\_COMB.seq.\*  
5: /cgn2.6/ptodata/1/ina/PCTUS\_COMB.seq.\*  
6: /cgn2.6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1183	81.6	1401	US-09-252-991A-7861	Sequence 7861, Ap
2	1167	80.5	1371	US-09-252-991A-7702	Sequence 7702, Ap
3	567	39.1	567	US-09-252-991A-7787	Sequence 7787, Ap
4	438	30.2	564	US-09-252-991A-7928	Sequence 7928, Ap
5	258	17.8	1359	US-09-252-991A-7624	Sequence 7624, Ap
6	160	11.0	360	US-09-252-991A-7552	Sequence 7552, Ap
7	120	8.3	1170	US-09-252-991A-7929	Sequence 7929, Ap
8	62	4.3	1287	US-09-252-991A-7860	Sequence 7860, Ap
9	36	2.5	1404	US-09-252-991A-7623	Sequence 7623, Ap
10	19	1.3	204	US-09-252-991A-9277	Sequence 9277, Ap
11	19	1.3	405	US-09-252-991A-1743	Sequence 1743, Ap
12	19	1.3	414	US-09-252-991A-12452	Sequence 12452, A
13	19	1.3	489	US-09-252-991A-8304	Sequence 8304, Ap
14	19	1.3	522	US-09-072-596-263	Sequence 263, App
15	19	1.3	723	US-09-252-991A-13195	Sequence 13195, A
16	19	1.3	873	US-09-252-991A-4459	Sequence 4459, Ap
17	19	1.3	888	US-09-252-991A-4147	Sequence 4147, Ap
18	19	1.3	930	US-09-252-991A-8815	Sequence 8815, Ap
19	19	1.3	942	US-09-252-991A-12760	Sequence 12760, A
20	19	1.3	966	US-09-252-991A-4666	Sequence 4666, Ap
21	19	1.3	1116	US-09-252-991A-10803	Sequence 10803, A
22	19	1.3	1206	US-09-252-991A-10627	Sequence 10627, A
23	19	1.3	1275	US-09-252-991A-10722	Sequence 10722, A
24	19	1.3	1641	US-09-252-991A-2066	Sequence 2066, Ap
25	19	1.3	1764	US-09-252-991A-16183	Sequence 16183, A
26	19	1.3	1770	US-09-252-991A-16408	Sequence 16408, A
27	19	1.3	1806	US-09-252-991A-15978	Sequence 15978, A

c	28	19	1.3	2112	4	US-09-252-991A-1820	Sequence 1820, Ap
	29	19	1.3	2283	4	US-09-252-991A-2140	Sequence 2140, Ap
	30	19	1.3	2932	4	US-09-016-434-1419	Sequence 1419, Ap
	31	19	1.3	4143	4	US-09-252-991A-9151	Sequence 9151, Ap
c	32	19	1.3	4212	4	US-09-252-991A-8929	Sequence 8929, Ap
c	33	19	1.3	14272	4	US-09-516-914-23	Sequence 23, Appl
	34	19	1.3	4403765	3	US-09-103-840A-2	Sequence 2, Appl
	35	19	1.3	4411529	3	US-09-103-840A-1	Sequence 1, Appl
	36	18	1.2	603	4	US-09-252-991A-1076	Sequence 1076, Ap
	37	18	1.2	792	4	US-09-252-991A-7005	Sequence 7005, Ap
	38	18	1.2	1308	4	US-09-501-115-31	Sequence 31, Appl
	39	18	1.2	1383	4	US-09-252-991A-16519	Sequence 16519, A
c	40	18	1.2	1488	4	US-09-252-991A-5798	Sequence 5798, Ap
c	41	18	1.2	1683	4	US-09-252-991A-10216	Sequence 10216, A
	42	18	1.2	1818	4	US-09-252-991A-9837	Sequence 9837, Ap
	43	18	1.2	1842	4	US-09-252-991A-5853	Sequence 5853, Ap
	44	18	1.2	1893	4	US-09-252-991A-9753	Sequence 9753, Ap
c	45	18	1.2	2028	4	US-09-252-991A-7091	Sequence 7091, Ap

ALIGNMENTS

RESULT 1  
US-09-252-991A-7861/c  
; Sequence 7861, Application US/09252991A.  
; Patent No. 6551795

GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; FILE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 7861

; LENGTH: 1401

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-7861

Query Match 81.6%; Score 1183; DB 4; Length 1401;

Best Local Similarity 99.7%; Pred. No. 0;

Matches 1383; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 GTGCTGATCGGCTCGCCACCTTGAAGCTGCGTTGAGGACGAGAGAGATGAGCTGAT 61

DB 1387 GTGCTGATCGGCTCGCCACCTTGAAGCTGCGTTGAGGACGAGAGATGAGCTGAT 1328

QY 62 CGCCTCGGACCACTTCCGCGATCGTTGCGGCTCGGCAAGAGGCGATGCTCTGCTGCG 121

DB 1327 CGCCTCGGACCACTTCCGCGATCGTTGCGGCTCGGCAAGAGGCGATGCTCTGCTGCG 1268

QY 122 CTACCTGGCGCGCGCGCTTGCCTTTGCGCGTGGTGCATACCGAGAGAACCCGCCGGA 181

DB 1267 CTACCTGGCGCGCGCGCGCTTGCCTTTGCGCGTGGTGCATACCGAGAGAACCCGCCGGA 1208

QY 182 GCTGGCCACCTTGCCTGCGGAGTATCGCCAGGTGGAAGTTCGTTGGCGGAACTCGACGC 241

DB 1207 GCTGGCCACCTTGCCTGCGGAGTATCGCCAGGTGGAAGTTCGTTGGCGGAACTCGACGC 1148

QY 242 CGAGTTCCTCTGCTCGGCGCGGAACTATGTCAGCCCGGCTTGTCTCGCTGGGACCC 301

DB 1147 CGAGTTCCTCTGCTCGGCGCGGAACTATGTCAGCCCGGCTTGTCTCGCTGGGACCC 1088

QY 302 TCGCTGTGTACAGGCGCGCGGAAAGGCGGCGCATCTCCGTCGATCATCTCTTCGC 361

DB 1087 CGCGTGTGTACAGGCGCGCGGAAAGGCGGCGCATCTCCGTCGATCATCTCTTCGC 1028

```
QY 362 CCGGAGGCGAAGCCCGGATGTCGCGCATCACCGTTCCAAACGCGAAGACACCGTGAC 421
DB 1027 CCGGAGGCGAAGCCCGGATGTCGCGCATCACCGTTCCAAACGCGAAGACACCGTGAC 968
QY 422 CACCTGTGTGGCAAAATGGCGGTGGCGGACAAAGCGTCTGCCGTGCGCGCAACCT 481
DB 967 CACCTGTGTGGCAAAATGGCGGTGGCGGACAAAGCGTCTGCCGTGCGCGCAACCT 908
QY 482 CGGACCCCGCGCTCGACCTGCTGGCGGACGACATCGAGCTGTACGTGTGGAGCTGTC 541
DB 907 CGGACCCCGCGCTCGACCTGCTGGCGGACGACATCGAGCTGTACGTGTGGAGCTGTC 848
QY 542 GAGCTTCAGCTGGAACCTGCGATCGCTCAACGCGAGGTGCGCACCGCTGCTGAACGT 601
DB 847 GAGCTTCAGCTGGAACCTGCGATCGCTCAACGCGAGGTGCGCACCGCTGCTGAACGT 788
QY 602 CAGCGAAGACCATATGATCGCTACGAGCGCATGGCTGACTACCACTGGCCAAAGCACCG 661
DB 787 CAGCGAAGACCATATGATCGCTACGAGCGCATGGCTGACTACCACTGGCCAAAGCACCG 728
QY 662 GATCTTCGCGGTGCCCGCAGGTGCTGGTGAATCGCGCGCATGCCCTGACCCGACCGCT 721
DB 727 GATCTTCGCGGTGCCCGCAGGTGCTGGTGAATCGCGCGCATGCCCTGACCCGACCGCT 668
QY 722 GATCGCGATACCGTGCCTGCTGCTGCTGCGCTGAACAGCCGACCTTCAAGGCTTT 781
DB 667 GATCGCGATACCGTGCCTGCTGCTGCTGCGCTGAACAGCCGACCTTCAAGGCTTT 608
QY 782 CGGCTGATCGAGAAAGCGCCAGAGTGGCTGGCTTCCAGTTCACAAAGCTGCTGCC 841
DB 607 CGGCTGATCGAGAAAGCGCCAGAGTGGCTGGCTTCCAGTTCACAAAGCTGCTGCC 548
QY 842 GGTGGGCACTGAAGATCCGTGGCGGCCCAACATATTCCAACGCGCTCGCGCGCTGGC 901
DB 547 GGTGGGCACTGAAGATCCGTGGCGGCCCAACATATTCCAACGCGCTCGCGCGCTGGC 488
QY 902 GCTGGGCACTGCGCTGCGCTGCTGCGAGCCCATGCTCGGCGCGTGAAGGCGTTTC 961
DB 487 GCTGGGCACTGCGCTGCGCTGCTGCGAGCCCATGCTCGGCGCGTGAAGGCGTTTC 428
QY 962 CGGCTTGGCTCATCGCTGCGAGTGGGTACGCGAGCGCGAGGGCTGAGCTACTACGACA 1021
DB 427 CGGCTTGGCTCATCGCTGCGAGTGGGTACGCGAGCGCGAGGGCTGAGCTACTACGACA 368
QY 1022 TTCAAAGGCCAACCACTCGCGCGCGCGCTGGCGCGATCGAGGGGCTGGGTGCGCGACAT 1081
DB 367 TTCAAAGGCCAACCACTCGCGCGCGCGCTGGCGCGATCGAGGGGCTGGGTGCGCGACAT 308
QY 1082 CGAGGCAAGCTGGTGTGCTGCTCGCGCGCGAGAGCGCAAGGCGCCGATTTCCATGACCT 1141
DB 307 CGAGGCAAGCTGGTGTGCTGCTCGCGCGCGAGAGCGCAAGGCGCCGATTTCCATGACCT 248
QY 1142 CGCGAGCGGCTCGCGCGCTTCTGCGCGCGGTGGTACTGCTGGCGGTGACGCGCGGCT 1201
DB 247 CGCGAGCGGCTCGCGCGCTTCTGCGCGCGGTGGTACTGCTGGCGGTGACGCGCGGCT 188
QY 1202 GATTGCCAGGCACTGGGCAACCGGTACCGCTGCTGCGCGGTGCGCAACGCTGGACGAAGC 1261
DB 187 GATTGCCAGGCACTGGGCAACCGGTACCGCTGCTGCGCGGTGCGCAACGCTGGACGAAGC 128
QY 1262 AGTCCGCGAGCGCGCGAGCTGCGCGCGAGGCGATGCGGTGCTGTCGCGCGGCTG 1321
DB 127 AGTCCGCGAGCGCGCGAGCTGCGCGCGAGGCGATGCGGTGCTGTCGCGCGGCTG 68
QY 1322 CGCGAGCGCTGGACATGTTCAAGAACTTCGAAGAACGCGGACGCTGTTCCGCAAGCGCT 1381
DB 67 CGCGAGCGCTGGACATGTTCAAGAACTTCGAAGAACGCGGACGCTGTTCCGCAAGCGCT 8
QY 1382 AGAGGAG 1388
DB 7 AGAGGAG 1
```

```
RESULT 2
US-09-252-991A-7702
; Sequence 7702, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 7702
; LENGTH: 1371
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-7702

Query Match      80.5%; Score 1167; DB 4; Length 1371;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1367; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 27 AGCTGCGCTTGAGGACGAAGAGAGCATGAGCTGATCGCTCCGACCACTTCCGCATCGTT 86
DB 1 AGCTGCGCTTGAGGACGAAGAGAGCATGAGCTGATCGCTCCGACCACTTCCGCATCGTT 60
QY 87 GTCGCGCTCGGCAAGAGCGCATGTCCTGTGTGCGTACCTTGGCGCGCGCGGCTTGCCT 146
DB 61 GTCGCGCTCGGCAAGAGCGCATGTCCTGTGTGCGTACCTTGGCGCGCGCGGCTTGCCT 120
QY 147 TTGCGCGTGTGCTATACCCGAGAGAACCCGCGGAGCTGGCCACCTTGGCGCGCGAGTAT 206
DB 121 TTGCGCGTGTGCTATACCCGAGAGAACCCGCGGAGCTGGCCACCTTGGCGCGCGAGTAT 180
QY 207 CCGCAGGTGGAAGTTCGCTTGGCGGAACTCGACCGGAGTTCCTCTGCTCGCGCGCGGAA 266
DB 181 CCGCAGGTGGAAGTTCGCTTGGCGGAACTCGACCGGAGTTCCTCTGCTCGCGCGCGGAA 240
QY 267 CTCTATGTCAGCCCCGGCTTGTGCTGCGCACCCCTTGGCTGTGTGCTGCTGCTGCTGCTG 326
DB 241 CTCTATGTCAGCCCCGGCTTGTGCTGCGCACCCCTTGTGCTGCTGCTGCTGCTGCTGCTG 300
QY 327 GCGGTGCGCATCTCCGCTGACATGCTCTTCCGCGCGGAGCGAAGGCCCGCATGCTC 386
DB 301 GGTGTGCGCATCTCCGCTGACATGCTCTTCCGCGCGGAGCGAAGGCCCGCATGCTC 360
QY 387 GCCATCACCGGTTCCAAACGGAAGACACCGTGAACACCTGCTGGCGCGGAAATGGCGGTG 446
DB 361 GCCATCACCGGTTCCAAACGGAAGACACCGTGAACACCTGCTGGCGCGGAAATGGCGGTG 420
QY 447 GCGCGGACAAAGCGTGTGCGCGTGGCGGCAACCTCGCGACCCCGCGCGCTGCGACCTGCTG 506
DB 421 GCGCGGACAAAGCGTGTGCGCGTGGCGGCAACCTCGCGACCCCGCGCGCTGCGACCTGCTG 480
QY 507 GCGCGACACATCGAGCTGTACGTGTGAGCTGTGAGCTTCCAGCTTGAACCTTGCAT 566
DB 481 GCGCGACACATCGAGCTGTACGTGTGAGCTGTGAGCTTCCAGCTTGAACCTTGCAT 540
QY 567 CGGCTCAACGCCGAGGTGGCGACCGTGTGAAGCTCAGCAAGACCAATATGGATGCTGCTAC 626
DB 541 CGGCTCAACGCCGAGGTGGCGACCGTGTGAAGCTCAGCAAGACCAATATGGATGCTGCTAC 600
QY 627 GAGCGCATGCTGACTACCACTGGCGCAAGACCGGATCTTCCGCGGTGCGCGCGAGTGC 686
DB 601 GAGCGCATGCTGACTACCACTGGCGCAAGACCGGATCTTCCGCGGTGCGCGCGAGTGC 660
QY 687 GTGTTGAATCGCGCGGATGCCCTGACCCGACCGCTGATCGCGCATACCGTGCCTGCTGG 746
DB 661 GTGTTGAATCGCGCGGATGCCCTGACCCGACCGCTGATCGCGCATACCGTGCCTGCTGG 720
```



QY 747 TCGTTCGGCTGAACAGCCGGACTTCAAGGCTTTCCGCTGATCGAGGAGACGGCCAG 806  
DB 721 TCGTTCGGCTGAACAGCCGGACTTCAAGGCTTTCCGCTGATCGAGGAGACGGCCAG 780  
QY 807 AAGTGGCTGGCGTTCCAGTTCGACAAAGCTGCTGCCGGTTGGCAAGCTGAAGATCCCGTGGC 866  
DB 781 AAGTGGCTGGCGTTCCAGTTCGACAAAGCTGCTGCCGGTTGGCAAGCTGAAGATCCCGTGGC 840  
QY 867 GCCACAACTATTCACAGCGCTCGCCGCTGCGCTGGCTGGCCATCGGTCGGCTCGC 926  
DB 841 GCCACAACTATTCACAGCGCTCGCCGCTGCGCTGGCTGGCCATCGGTCGGCTCGC 900  
QY 927 TTCGACGCCATGCTCGCGCGCTGAAGCGCTTTTCGGCGCTGGCTCATCTGCTGCAAGTGG 985  
DB 901 TTCGACGCCATGCTCGCGCGCTGAAGCGCTTTTCGGCGCTGGCTCATCTGCTGCAAGTGG 960  
QY 987 GTACCGAGCGGCGAGCGCTGAGCTACTACGACGATTCACAGGCGCCACCAAGCTCGCGGCC 1046  
DB 961 GTACCGAGCGGCGAGCGCTGAGCTACTACGACGATTCACAGGCGCCACCAAGCTCGCGGCC 1020  
QY 1047 GCCCTGCGCGGATCGAGGGCTGGGTGCGACATCGACGCAAGCTGGTGTCTGCTGCC 1106  
DB 1021 GCCCTGCGCGGATCGAGGGCTGGGTGCGACATCGACGCAAGCTGGTGTCTGCTGCC 1080  
QY 1107 GCGGAGACGCGCAAGGCGCGGATTTCCATGACCTGCGCGAGCGGTCGCGCGCTTCTGC 1166  
DB 1081 GCGGAGACGCGCAAGGCGCGGATTTCCATGACCTGCGCGAGCGGTCGCGCGCTTCTGC 1140  
QY 1167 CGCGCGTGTACTGCTTGGCCGTGACCGCGGCTGATTCGCCAGGCACTGGGCAAGCG 1226  
DB 1141 CGCGCGTGTACTGCTTGGCCGTGACCGCGGCTGATTCGCCAGGCACTGGGCAAGCG 1200  
QY 1227 GTACCGCTGTGCGCTCGCAACGCTGGCAAGCACTGCGGCAAGCGCGCGAGCTGGCC 1286  
DB 1201 GTACCGCTGTGCGCTCGCAACGCTGGCAAGCACTGCGGCAAGCGCGCGAGCTGGCC 1260  
QY 1287 CGCAAGCGATGCGGTGCTGTTGTCGCCGGCTGCGGAGCGCTTGGACATGTTCAAGAAC 1346  
DB 1261 CGCAAGCGATGCGGTGCTGTTGTCGCCGGCTGCGGAGCGCTTGGACATGTTCAAGAAC 1320  
QY 1347 TTCGAAGACCGGACGCGCTTTCGCCAAGCGCTAGAGGAGCTAGCGTGA 1397  
DB 1321 TTCGAAGACCGGACGCGCTTTCGCCAAGCGCTAGAGGAGCTAGCGTGA 1371

## RESULT 3

US-09-252-991A-7787/c  
; Sequence 7787, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252.991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 7787  
; LENGTH: 567  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-7787

Query Match 39.1%; Score 567; DB 4; Length 567;  
Best Local Similarity 100.0%; Pred. No. 1e-251;  
Matches 567; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 772 TCAAGGCTTTTCGGCTGATCGAGGAGACGGCCAGAGTGGCTGGCGTTCCAGTTCGACA 831

DB 567 TCAAGGCTTTTCGGCTGATCGAGGAGACGGCCAGAGTGGCTGGCGTTCCAGTTCGACA 508  
QY 832 AGCTGCTGCCGGTTGGCGAACTGAAGATCCGTTGGCGCCCAACAACATATTCAACGCGCTCG 891  
DB 507 AGCTGCTGCCGGTTGGCGAACTGAAGATCCGTTGGCGCCCAACAACATATTCAACGCGCTCG 448  
QY 892 CCGCGCTGGCGCTGGCGCATCGGCTGCGCTGCGCTTCGAGCGGCATGCTCGGCGCGCTGA 951  
DB 447 CCGCGCTGGCGCTGGCGCATCGGCTGCGCTGCGCTTCGAGCCCATGCTCGGCGCGCTGA 388  
QY 952 AGCGTTTTCGGCGCTGGCTCATCGCTGCGAGTGGGTACGCGAGCGGAGCGGTGAGCT 1011  
DB 387 AGCGTTTTCGGCGCTGGCTCATCGCTGCGAGTGGGTACGCGAGCGGAGCGGTGAGCT 328  
QY 1012 ACTACGACGATTCACAGGCGCCCAACCTGCGCGCGCCCTGGCGGCATCGAGGGCGCTGG 1071  
DB 327 ACTACGACGATTCACAGGCGCCCAACCTGCGCGCGCCCTGGCGGCATCGAGGGCGCTGG 268  
QY 1072 GTGCCGACATCGAGCGCAAGCTGGTGTGCTGCTGCGCGCGGAGAGCGCAAGGCGCGCGATT 1131  
DB 267 GTGCCGACATCGAGCGCAAGCTGGTGTGCTGCTGCGCGCGGAGAGCGCAAGGCGCGCGATT 208  
QY 1132 TCATGACCTGCGGAGCGCGCTGCGCGCTTCTGCGCGCGGCGGTGCTACTGCTGGCGGTG 1191  
DB 207 TCCATGACCTGCGGAGCGCGCTGCGCGCTTCTGCGCGCGGCGGTGCTACTGCTGGCGGTG 148  
QY 1192 ACGCCGGCTGATTCGCCAGGCAAGCTGGTGGCAACGCGGTACCGCTGGTGGCTGCGCAACGC 1251  
DB 147 ACGCCGGCTGATTCGCCAGGCAAGCTGGTGGCAACGCGGTACCGCTGGTGGCTGCGCAACGC 88  
QY 1252 TGGACGAAGCAGTTCGCGAGCGCGCGAGCTGGCGCGCGAGCGATCGCGTGTGTTGT 1311  
DB 87 TGGACGAAGCAGTTCGCGAGCGCGCGAGCTGGCGCGCGAGCGATCGCGTGTGTTGT 28  
QY 1312 CGCGGCGCTGCGGAGCGCTGGACATGT 1338  
DB 27 CGCGGCGCTGCGGAGCGCTGGACATGT 1

## RESULT 4

US-09-252-991A-7928/c  
; Sequence 7928, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252.991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 7928  
; LENGTH: 564  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-7928

Query Match 30.2%; Score 438; DB 4; Length 564;  
Best Local Similarity 100.0%; Pred. No. 2.5e-192;  
Matches 438; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1013 CTACGACGATTCGAAGGCCACCAAGCTGGCGCGCGCCCTGGCGCGCATCGAGGGCGCTGG 1072  
DB 564 CTACGACGATTCGAAGGCCACCAAGCTGGCGCGCGCCCTGGCGCGCATCGAGGGCGCTGG 505  
QY 1073 TGCCGACATCGACGGCAAGCTGGTGTGCTGCGCGCGGAGACGCGCAAGGCGCGCGATT 1132  
DB 504 TGCCGACATCGACGGCAAGCTGGTGTGCTGCTGCGCGCGGAGACGCGCAAGGCGCGCGATT 445

```

QY 1133 CCATGACCTGCGGAGCGGTGCGCGCTTCTCCGCGGGTGGTACTGCTGGCGGTGA 1192
Db 444 CCATGACCTGCGGAGCGGTGCGCGCTTCTCCGCGGGTGGTACTGCTGGCGGTGA 385
QY 1193 CGCCGGCTGATTTGCCAGGCACTGGGCAACGCGGTACCGCTGGTGGCGGTGCGCAACGCT 1252
Db 384 CGCCGGCTGATTTGCCAGGCACTGGGCAACGCGGTACCGCTGGTGGCGGTGCGCAACGCT 325
QY 1253 GGACGAGCAGTCCGCGAGCGCGCGAGCTGGCCCGGGAAGGGGATGCGGTGTTGTC 1312
Db 324 GGACGAGCAGTCCGCGAGCGCGCGAGCTGGCCCGGGAAGGGGATGCGGTGTTGTC 265
QY 1313 GCGCGCTCGCGGAGCTGGACATGTTCAAGAACTTCGAAGAACGCGGACGCTGTTCCG 1372
Db 264 GCGCGCTCGCGGAGCTGGACATGTTCAAGAACTTCGAAGAACGCGGACGCTGTTCCG 205
QY 1373 CAAAGCCGTAGAGGAGCTAGCGTGTGTCGGTGTTCGCGCCCTTCCCGTCGCGGTG 1432
Db 204 CAAAGCCGTAGAGGAGCTAGCGTGTGTCGGTGTTCGCGCCCTTCCCGTCGCGGTG 145
QY 1433 TTGAGCGCGCAGGCATC 1450
Db 144 TTGAGCGCGCAGGCATC 127

```

```

RESULT 5
US-09-252-991A-7624
Sequence 7624, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 7624
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-7624

```

```

Query Match 17.8%; Score 258; DB 4; Length 1359;
Best Local Similarity 100.0%; Pred. No. 1.6e-109;
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1193 CGCCGGCTGATTTGCCAGGCACTGGGCAACGCGGTACCGCTGGTGGCGGTGCGCAACGCT 1252
Db 1 CGCCGGCTGATTTGCCAGGCACTGGGCAACGCGGTACCGCTGGTGGCGGTGCGCAACGCT 60
QY 1253 GGACGAGCAGTCCGCGAGCGCGCGAGCTGGCCCGGGAAGGGGATGCGGTGTTGTC 1312
Db 61 GGACGAGCAGTCCGCGAGCGCGCGAGCTGGCCCGGGAAGGGGATGCGGTGTTGTC 120
QY 1313 GCGCGCTCGCGGAGCTGGACATGTTCAAGAACTTCGAAGAACGCGGACGCTTTCGC 1372
Db 121 GCGCGCTCGCGGAGCTGGACATGTTCAAGAACTTCGAAGAACGCGGACGCTTTCGC 180
QY 1373 CAAAGCCGTAGAGGAGCTAGCGTGTGTCGGTGTTCGCGCCCTTCCCGTCGCGGTG 1432
Db 181 CAAAGCCGTAGAGGAGCTAGCGTGTGTCGGTGTTCGCGCCCTTCCCGTCGCGGTG 240
QY 1433 TTGAGCGCGCAGGCATC 1450
Db 241 TTGAGCGCGCAGGCATC 258

```

```

RESULT 6
US-09-252-991A-7552
Sequence 7552, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 7552
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-7552

```

```

Query Match 11.0%; Score 160; DB 4; Length 360;
Best Local Similarity 99.5%; Pred. No. 2.3e-64;
Matches 210; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 61 TCGCCTCGGACCACTTCCGCATCGTTGCGGCTCGGCAAGAGCGGCATGTCCTGGTGC 120
Db 1 TCGCCTCGGACCACTTCCGCATCGTTGCGGCTCGGCAAGAGCGGCATGTCCTGGTGC 60
QY 121 GCTACCTGGCGCGCGCGCTTTCCTTTGCGCGTGTGATACCCGAGAGAACCCGCGG 180
Db 61 GCTACCTGGCGCGCGCGCTTTCCTTTGCGCGTGTGATACCCGAGAGAACCCGCGG 120
QY 181 AGCTGGCACCTCGTGGCCAGTATCCGAGGTGGAAGTGGTTCGCGGCAACTCGACG 240
Db 121 AGCTGGCACCTCGTGGCCAGTATCCGAGGTGGAAGTGGTTCGCGGCAACTCGACG 180
QY 241 CCGAGTTCCTCTGCTCCGCGCGCGCAACTCTA 271
Db 181 CCGAGTTCCTCTGCTCCGCGCGCGCAACTCTA 211

```

```

RESULT 7
US-09-252-991A-7929/c
Sequence 7929, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 7929
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-7929

```

```

Query Match 8.3%; Score 120; DB 4; Length 1170;
Best Local Similarity 100.0%; Pred. No. 5.5e-46;
Matches 120; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GTGCTGATCGGCTCGCCACCTTGAAGCTGCTTGGAGGACGAGAGCATGAGCCTGAT 61
Db 120 GTGCTGATCGGCTCGCCACCTTGAAGCTGCTTGGAGGACGAGAGCATGAGCCTGAT 61

```





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 14, 2003, 09:25:27 ; Search time 91 Seconds  
(without alignments)  
2172.963 Million cell updates/sec

Title: US-09-701-229-2

Perfect score: 448

Sequence: 1 MSLIAHDHFRIIVVLGKSGM.....MFKNFERGLFAKAVEELA 448

Scoring table:

OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 569978 seqs, 220691566 residues

Word size: 1

Total number of hits satisfying chosen parameters: 712408

Minimum DB seq length: 25

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DRV=xlp  
-Q=/cgn2\_1/USPTO.spool/US09701229/runat\_09082003\_161237\_5050/app\_query.fasta\_1.647  
-DB=Issued\_Patents\_NA -QFWT=fastap -SUFFIX=olip2n.rni -MINMATCH=0.1 -LOOPL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR\_SCORE=quality -THR\_MIN=1 -ALIGN=15 -MODE=LOCAL  
-OUTFWT=pto -NORM=ext -HEAPSIZ=500 -MINLEN=25 -MAXLEN=2000000000  
-USER=US09701229 @CEN\_1\_1\_85 @runat\_09082003\_161237\_5050 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6  
-FGAPEXT=7 -FGAPOP=60 -YGAPOP=60 -DELOP=6 -DELEXT=7

Database :

Issued\_Patents\_NA.\*  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/PTUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	448	100.0	1371	4	US-09-252-991A-7702
C 2	446	99.6	1401	4	US-09-252-991A-7861
C 3	188	42.0	567	4	US-09-252-991A-7787
C 4	127	28.3	564	4	US-09-252-991A-7928
C 5	119	26.6	360	4	US-09-252-991A-7552
C 6	67	15.0	1359	4	US-09-252-991A-7624
C 7	23	5.1	1170	4	US-09-252-991A-7929
C 8	17	3.8	1416	4	US-09-328-352-1282
C 9	14	3.1	1830121	4	US-09-557-884-1
C 10	14	3.1	1830121	4	US-09-643-990A-1
C 11	9	2.0	297	4	US-09-313-294A-6271
C 12	9	2.0	298	4	US-09-313-294A-5881

13	9	2.0	717	4	US-08-936-165A-78	Sequence 78, Appl
14	9	2.0	1242	3	US-09-147-928-1	Sequence 1, Appl
15	9	2.0	1372	4	US-09-530-836-1	Sequence 1, Appl
16	9	2.0	1374	4	US-09-134-001C-1001	Sequence 1001, Ap
17	9	2.0	2423	3	US-08-714-918-86	Sequence 86, Appl
18	9	2.0	2423	3	US-09-265-315-86	Sequence 86, Appl
19	9	2.0	2423	3	US-09-265-315-86	Sequence 86, Appl
20	9	2.0	2423	3	US-09-266-417-86	Sequence 86, Appl
C 21	8	1.8	506	2	US-08-799-173A-15	Sequence 15, Appl
C 22	8	1.8	529	4	US-09-109-204-28	Sequence 28, Appl
C 23	8	1.8	529	4	US-09-490-032-28	Sequence 28, Appl
C 24	8	1.8	655	3	US-09-109-204-27	Sequence 27, Appl
C 25	8	1.8	655	3	US-09-490-032-27	Sequence 27, Appl
C 26	8	1.8	855	4	US-09-252-991A-13763	Sequence 13763, A
C 27	8	1.8	873	4	US-09-252-991A-13662	Sequence 13662, A
C 28	8	1.8	963	4	US-09-252-991A-6979	Sequence 6979, Ap
C 29	8	1.8	990	4	US-09-252-991A-13865	Sequence 13865, A
C 30	8	1.8	1105	2	US-08-799-173A-1	Sequence 1, Appl
C 31	8	1.8	1131	4	US-09-252-991A-7006	Sequence 7006, Ap
C 32	8	1.8	1164	4	US-09-252-991A-1310	Sequence 1310, Ap
C 33	8	1.8	1326	4	US-09-252-991A-1156	Sequence 1156, Ap
C 34	8	1.8	1419	4	US-09-252-991A-1246	Sequence 1246, Ap
C 35	8	1.8	1428	4	US-09-252-991A-13671	Sequence 13671, A
C 36	8	1.8	1557	4	US-09-252-991A-7033	Sequence 7033, Ap
C 37	8	1.8	1779	3	US-09-371-696-1	Sequence 1, Appl
C 38	8	1.8	1886	4	US-09-484-970B-127	Sequence 127, Ap
C 39	8	1.8	2028	4	US-09-252-991A-7091	Sequence 7091, Ap
C 40	8	1.8	2224	3	US-09-109-204-6	Sequence 6, Appl
C 41	8	1.8	2224	3	US-09-490-032-6	Sequence 6, Appl
C 42	8	1.8	2360	1	US-08-466-390-1	Sequence 1, Appl
C 43	8	1.8	2360	1	US-08-470-950-1	Sequence 1, Appl
C 44	8	1.8	2360	1	US-08-467-781-1	Sequence 1, Appl
C 45	8	1.8	2360	1	US-08-195-487-1	Sequence 1, Appl

#### ALIGNMENTS

##### RESULT 1

US-09-252-991A-7702  
; Sequence 7702, Application US/09252991A  
; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 7702

; LENGTH: 1371

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-7702

Alignment Scores:

Pred. No.: 0 Length: 1371  
Score: 448.00 Matches: 448  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 4 Gaps: 0

US-09-701-229-2 (1-448) x US-09-252-991A-7702 (1-1371)

Qy 1 MetSerLeuIleAlaSerAspHisPheArgIleValValGlyLeuGlyLysSerGlyMet 20

|||||  
25 ATGAGCGCTGATCGCCTCCGACCACTCCGCGCATGTTCGCGTTCGCGAAGCGCATG 84

QY 21 SerLeuValArgTyrLeuAlaArgArgGlyLeuProPheAlaValValAspThrArgGlu 40  
DB 85 TCCCTGGTGGCTACTGCGCGCGCGCTTGCCTTGCCTGGTATACCGAGAG 144  
QY 41 AsnProGluLeuAlaThrLeuArgAlaGlnTyrProGlnValGluValArgCysGly 60  
DB 145 AACCCGCGGAGCTGCCACCTGCTGCGCCAGTATCCGACGAGTGGAGTGGTGGCGC 204  
QY 61 GluLeuAspAlaGluPheLeuCysSerAlaArgGluLeuTyrValSerProGlyLeuSer 80  
DB 205 GAACTCGAGCGCGAGTCTCTGCTCCGCCCGGAACCTAGTCAGCCCGCGCTTGGC 264  
QY 81 LeuArgThrProAlaLeuValGlnAlaAlaLysGlyValArgIleSerGlyAspIle 100  
DB 265 CTGCGCACCCCGCGCTGCTACAGCGCGCGGAAAGGTGTCGCATCTCCGGTGACATC 324  
QY 101 AspLeuPheAlaArgGluAlaLysAlaProIleValAlaIleThrGlySerAsnAlaLys 120  
DB 325 GATCTCTTCGCGCGGAGCGGAAGCCCGCATGCTGCGCATCACCGGTTCACACGGAAG 384  
QY 121 SerThrValThrThrLeuValGlyGluMetAlaValAlaAlaAspLysArgValAlaVal 140  
DB 385 AGCACCGTGACACCTGCTGGCGGAATGCGGTGGCGCGGACAAAGCTGTCGCCCTC 444  
QY 141 GlyGlyAsnLeuGlyThrProAlaLeuAspLeuLeuAlaAspIleGluLeuTyrVal 160  
DB 445 GCGCGCAACCTCGGACCCCGCGCTCGACCTGCTGGCGGAGACATCGAGCTGTACCTG 504  
QY 161 LeuGluLeuSerSerPheGlnLeuGlnThrCysAspArgLeuAsnAlaGluValAlaThr 180  
DB 505 TTGGAGCTGTCGAGCTTCAGCTGGAACCTGCGATCCCTCAACCGCGAGTGGCGACC 564  
QY 181 ValLeuAsnValSerGluAspHisMetAspArgTyrAspGlyMetAlaAspTyrHisLeu 200  
DB 565 GTGCTGAACCTCAGCGAAGACCATATGATCGCTACGACGATGCTGCTACACCTG 624  
QY 201 AlaLysHisArgIlePheArgGlyAlaArgGlnValValAlaAsnArgAlaAspAlaLeu 220  
DB 625 GCCAAGCAGCGATCTTCGCGGTGCGCGGAGTGGTGGTGAATCGCGCGCGATGCCCTG 684  
QY 221 ThrArgProLeuIleAlaAspThrValProCysTyrPheGlyLeuAsnLysProAsp 240  
DB 685 ACCCGACCGCTGATCGCGGATACCTGCGTGGTGGTGGTGGTGGTGGTGGTGGTGG 744  
QY 241 PheLysAlaPheGlyLeuIleGluAlaAspGlyGlnLysTyrLeuAlaPheGlnPheAsp 260  
DB 745 TTCAGGCTTTCGCGCTGATCGAGAGACGCGGAGAGTGGTGGTGGTGGTGGTGGTGG 804  
QY 261 LysLeuLeuProValGlyGluLeuLysIleArgGlyAlaHisAsnTyrSerAsnAlaLeu 280  
DB 805 AAGCTGCTGCGCGTGGCGAATGAAGATCCGTCGCGCGGCGGCGGCGGCGGCGGCGG 864  
QY 281 AlaAlaLeuAlaLeuGlyHisAlaValGlyLeuProPheAspAlaMetLeuGlyAlaLeu 300  
DB 865 GCGCGCTGCGCTGGCGCATCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCG 924  
QY 301 LysAlaPheSerGlyLeuAlaHisArgCysGlnTyrValArgGluArgGlnGlyValSer 320  
DB 925 AAGCGCTTTCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 984  
QY 321 TyrTyrAspSerLysAlaThrAsnValGlyAlaAlaLeuAlaAlaIleGluGlyLeu 340  
DB 985 TACTACGACGATTCAGCGCCACCACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1044  
QY 341 GlyAlaAspIleAspGlyLysLeuValLeuLeuAlaGlyGlyAspGlyLysGlyAlaAsp 360  
DB 1045 GGTGCGGACATCGAGCGCAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1104  
QY 361 PheHisAspLeuArgGluProValAlaArgPheCysArgAlaValValLeuLeuGlyArg 380  
DB 1105 TTCCATGACCTGCGCGAGCGCGCTTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 1164  
QY 381 AspAlaGlyLeuIleAlaGlnAlaLeuGlyAsnAlaValProLeuValArgValAlaThr 400

DB 1165 GACGCGGCGCTGATTCGCCAGCACTGGGCAACGGGTACCGCTGCTGGCTGCAACG 1224  
QY 401 LeuAspGluAlaValArgGlnAlaAlaGlnLeuAlaArgGluGlyAspAlaValLeuLeu 420  
DB 1225 CTGGACGAAGCAGTCCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1284  
QY 421 SerProAlaCysAlaSerLeuAspMetPheLysAsnPheGluGluArgGlyArgLeuPhe 440  
DB 1285 TCGCGGCGCTGCGCGAGCTGGACATGTTCAAGAACTTCGAAGAACGCGGCGCTGTC 1344  
QY 441 AlaLysAlaValGluGluLeuAla 448  
DB 1345 GCCAAAGCGTAGAGGAGTAGCG 1368  
RESULT 2  
US-09-252-991A-7861/c  
; Sequence 7861, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252.991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 7861  
; LENGTH: 1401  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-7861  
Alignment Scores:  
Pred. No.: 0 Length: 1401  
Score: 446.00 Matches: 446  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 99.55% Indels: 0  
DB: 4 Gaps: 0  
US-09-701-229-2 (1-448) x US-09-252-991A-7861 (1-1401)  
QY 1 MetSerLeuIleAlaSerAspHisPheArgIleValValGlyLeuGlyLysSerGlyMet 20  
DB 1338 ATGAGCTGATCGCTCGACCACTTCGCGCATCTGTCGCGCTCGGCAAGCGCGCATG 1279  
QY 21 SerLeuValArgTyrLeuAlaArgArgGlyLeuProPheAlaValValAspThrArgGlu 40  
DB 1278 TCCCTGGTGGCTACTGCGCGCGCGCGCTTGCCTTTCGCGCTTGCATACCGAGAG 1219  
QY 41 AsnProGluLeuAlaThrLeuArgAlaGlnTyrProGlnValGluValArgCysGly 60  
DB 1218 AACCCGCGGAGCTGGCGACCCCTGCTGCGCGAGTATCCGCGAGTGGAGTGGCTGGCG 1159  
QY 61 GluLeuAspAlaGluPheLeuCysSerAlaArgGluLeuTyrValSerProGlyLeuSer 80  
DB 1158 GAATCGACGCGAGTTCCTCTGCTCGCGCGGCACTTACGTACGCGCGCGCTGCTGCG 1099  
QY 81 LeuArgThrProAlaLeuValGlnAlaAlaLysGlyValArgIleSerGlyAspIle 100  
DB 1098 CTGCGCACCCCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1039  
QY 101 AspLeuPheAlaArgGluAlaLysAlaProIleValAlaIleThrGlySerAsnAlaLys 120  
DB 1038 GATCTCTTCGCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 979  
QY 121 SerThrValThrThrLeuValGlyGluMetAlaValAlaAspLysArgValAlaVal 140



Db 978 AGCACCGTACCCCTGGTGGCGAAATGGCGGTGCGCGGACGAAGCGTGTGCGCGTC 919  
Qy 141 GlyGlyAsnLeuGlyThrProAlaLeuAspLeuAlaAspAspIleGluLeuTyrVal 160  
Db 918 GCGGCAACCTCGGCACCCCGCGCTGACCTGCTGCGCGACACATCGAGCTGTACGTG 859  
Qy 161 LeuGluLeuSerPheGlnLeuGluThrCysAspArgLeuAsnAlaGluValAlaThr 180  
Db 858 TTGGAGCTGTGAGCTTCCAGCTTGAACCTGCGATCGCTCAACGCGGAGTGGCGACC 799  
Qy 181 ValLeuAsnValSerGluAspHisMetAspArgTyrAspGlyMetAlaAspTyrHisLeu 200  
Db 798 GTGCTGAACGTACGGAAGACCATATGATGATCGGTACGCGCATGGCTACCACTG 739  
Qy 201 AlalysHisArgIlePheArgGlyAlaArgGlnValValValAsnArgAlaAspAlaLeu 220  
Db 738 GCCAAGCACCGGATCTCCGCGGTGCGCGCGCAGTCTGTGTAATCGCGCGATGCCCTG 679  
Qy 221 ThrArgProLeuIleAlaAspThrValProCysTyrPheGlyLeuAsnLysProAsp 240  
Db 678 ACCGACCGGTGATCGCGCATACCGTGCCTGCTGTGCTGCTGCTGCTGCTGCTGCTG 619  
Qy 241 PheLysAlaPheGlyLeuIleGluGluAspGlyGlnLysTrpLeuAlaPheGlnPheAsp 260  
Db 618 TTCAAGGCTTTCGCGCTGATCGAGGAAGACGCGCAGAGTGGCTGGCGTTCCAGTTCCGAC 559  
Qy 261 LysLeuLeuProValGlyGluLeuLysIleArgGlyAlaHisAsnTyrSerAsnAlaLeu 280  
Db 558 AAGCTGCTCGCGTTGCGGAATCAAGATCCGTGGCGCCACAACTATTCCACGCGCTC 499  
Qy 281 AlaAlaLeuAlaLeuGlyHisAlaValGlyLeuProPheAspAlaMetLeuGlyAlaLeu 300  
Db 498 GCGCGCTGCGGTGGCGCATCGGTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCG 439  
Qy 301 LysAlaPheSerGlyLeuAlaHisArgCysGlnTrpValArgGluArgGlnGlyValSer 320  
Db 438 AAGCGTTTTCGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTG 379  
Qy 321 TyrTyrAspSerLysAlaThrAsnValGlyAlaAlaAlaLeuAlaIleGluGlyLeu 340  
Db 378 TACTACGACGTTCCAGGCGCCACCAAGCTGCGCGCGCGCTGCGCGGATCGAGGGCTG 319  
Qy 341 GlyAlaAspIleAspGlyLysLeuValLeuAlaGlyLysAspGlyLysGlyAlaAsp 360  
Db 318 GGTGCCGACATCGACGCAAGCTGTGCTGCTGCGCGCGGAGCGCAAGGCGCGCAT 259  
Qy 361 PheHisAspLeuArgGluProValAlaArgPheCysArgAlaValValLeuLeuGlyArg 380  
Db 258 TTCCATGACCTGGCGAGCGCGTGGCGCTTCTGCGCGCGGTGCTGCTGCTGCTGCTG 199  
Qy 381 AspAlaGlyLeuIleAlaGlnAlaLeuGlyAsnAlaValProLeuValArgValAlaThr 400  
Db 198 GACGCGCGGTGATGCGCGAGGACTGCGGCAACGCGGTACCGTGTGCGCGCTGCGACG 139  
Qy 401 LeuAspGluAlaValArgGlnAlaAlaGluLeuAlaArgGluGlyAspAlaValLeuLeu 420  
Db 138 CTGGACGAAGCAGTCCGCGAGCGCGCGAGTGGCGCGGCAAGCGGATGCGTGTGTTG 79  
Qy 421 SerProLacysAlaSerLeuAspMetPheLysAsnPheGluArgGlyArgLeuPhe 440  
Db 78 TCGCGCGCTCGCGAGCGCTGGACATGTTCAAGAACTTCGAAGAACCGCGACGCTGTTT 19  
Qy 441 AlalysAlaValGluGlu 446  
Db 18 GCCAAACCGGTAGAGGAG 1

RESULT 3

US-09-252-991A-7787/c  
; Sequence 7787, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 7787  
; LENGTH: 567  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-7787

Alignment Scores:

Pred. No.: 1,73e-175 Length: 567  
Score: 188.00 Matches: 188  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 41.96% Indels: 0  
DB: 4 Gaps: 0

US-09-701-229-2 (1-448) x US-09-252-991A-7787 (1-567)

Qy 242 LysAlaPheGlyLeuIleGluGluAspGlyGlnLysTrpLeuAlaPheGlnPheAspLys 261  
Db 565 AAGGCTTCGCGCTGATCGAGGAAGACGCGCAGAGTGGCTGCGTTCACATTCGACAA 506  
Qy 262 LeuLeuProValGlyGluLeuLysIleArgGlyAlaHisAsnTyrSerAsnAlaLeuAla 281  
Db 505 CTGCTGCGCGTTGGCAACTGAAGATCCGTGGCGCCACAACTATTCCACGCGCTGCC 446  
Qy 282 AlaLeuAlaLeuGlyHisAlaValGlyLeuProPheAspAlaMetLeuGlyAlaLeuLys 301  
Db 445 GCCTGCGCTGGCGCATCGGTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTG 386  
Qy 302 AlaPheSerGlyLeuAlaHisArgCysGlnTrpValArgGluArgGlnGlyValSerTyr 321  
Db 385 GCGTTCGCGCTGCGCTCATCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCG 326  
Qy 322 TyrAspAspSerLysAlaThrAsnValGlyAlaAlaLeuAlaIleGluGlyLeuGly 341  
Db 325 TAGCAGATTCGAAGCGCAACACGTCGCGCGCGCTGCGCGCGCTGCGCGCGCTGCG 266  
Qy 342 AlaAspIleAspGlyLysLeuValLeuAlaGlyLysAspGlyLysGlyAlaAspPhe 361  
Db 265 GCGGACATCGACGCGAAGCTGGTGTGCTGCTGCGCGCGGAGACGCGCGCGCTGCT 206  
Qy 362 HisAspLeuArgGluProValAlaArgPheCysArgAlaValValLeuLeuGlyArgAsp 381  
Db 205 CATGACCTGCGCGAGCGCGTGGCTGCGCTGCTGCGCGCGCTGCTGCTGCTGCTGCT 146  
Qy 382 AlaGlyLeuIleAlaGlnAlaLeuGlyAsnAlaValProLeuValArgValAlaThrLeu 401  
Db 145 GCGCGGCTGATTCCCGAGGACCTGGGCAACGCGGTACCGCTGGTGGCGCTGCGAACG 86  
Qy 402 AspGluAlaValArgGlnAlaAlaGluLeuAlaArgGluGlyAspAlaValLeuLeuSer 421  
Db 85 GAGGAGCAGTTCGCGAGCGCGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTG 26  
Qy 422 ProAlaCysAlaSerLeuAspMet 429  
Db 25 CCGCGCTGCGCGAGCTGGACATG 2

RESULT 4

US-09-252-991A-7928/c  
; Sequence 7928, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

```

; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 7928
; LENGTH: 564
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-7928

```

Alignment Scores:	
Pred. No.:	1.26e-115
Score:	127.00
Length:	564
Matches:	127
Conservative:	0
Percent Similarity:	100.00%
Best Local Similarity:	100.00%
Mismatches:	0
Query Match:	28.35%
Indels:	4
Gaps:	0
DB:	0

US-09-701-229-2 (1-448) x US-09-252-991A-7928 (1-564)

322	TyrAspAspSerLysAlaThrAsnValGlyAlaAlaLeuAlaAlaIleGluGlyLeuGly	341
QY		
Dbb		
563	TACGACGATTCCAAGGCGCACCAAGCTCGCGCGCGCGCTGCGCGCATGAGGGGCTGGGT	504
QY		
Dbb		
342	AlaAspIleAspGlyLysLeuValLeuAlaAlaGlyAspGlyLysGlyAlaAspPhe	361
QY		
Dbb		
503	GCCGACATCGACGCAAGCTGGTGTCTCGCCGGCGGACAGCGCAAGGCGCCGATTTC	444
QY		
Dbb		
362	HisAspLeuArgGluProValAlaArgPheCysArgAlaValValLeuLeuGlyArgAsp	381
QY		
Dbb		
443	CATGACCTGCGGAGCGCGTCTGCGCGGGCGGTGTACTGTTGGCCGTGAC	384
QY		
Dbb		
382	AlaGlyLeuIleAlaGlnAlaLeuGlyAsnAlaValProLeuValArgValAlaThrLeu	401
QY		
Dbb		
383	CGCGGGTGATTGCCAGGCACATGGGCAACGCGGTACCGCTGGTGCCTGCAAGCGTG	324
QY		
Dbb		
402	AspGluAlaValArgGlnAlaAlaGluLeuAlaArgGluGlyAspAlaValLeuLeuSer	421
QY		
Dbb		
323	GACGAAGCAGTCCGGCAGGCGCGGAGCTGCGCCGCAAGGCGATCGGTGTGTGTGCG	264
QY		
Dbb		
422	ProAlaCysAlaSerLeuAspMetPheLysAsnPheGluArgGlyArgLeuPheAla	441
QY		
Dbb		
263	CCGCGCTCGCGGACCTTGGACATGTTCCAGAACTTCAGAAACCCGCGCCTTTCGCC	204
QY		
Dbb		
442	LysAlaValGluLeuAla	448
QY		
Dbb		
203	AAAGCCGTAGAGGACTAGCG	183
QY		
Dbb		

## RESULT 5

US-09-252-991A-7552  
; Sequence 7552, Application US/09252991A  
; Patent No. 6551795  
GENERAL INFORMATION:

APPLICANT: MARC J. RUDENIIFELD et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

Alignment Scores:  
Pred. No.:  
Score:  
Percent Similarity:  
Best Local Similarity:  
Query Match:  
DB:

US-09-701-229-2 (1-448) x US-09-252-991A-7552 (1-360)

QY	5	AlaSerAspHisPheArgIleValValGlyLeuGlyLysSerGlyMetSerLeuValArg	24
DB	3	GCCTCCGACCACTTCCGCATCGTTGTCGGCTCGGCAAGACGGCATGTCCCTGGTGGCC	62
QY	25	TyrLeuAlaArgArgGlyLeuProPheAlaValValAspThrArgGluAsnProProGlu	44
DB	63	TACCTGGCGCGCCGCGCTTGCCTTTCGCCGTTGTCGATACCCGAGAGAACCCGCCGGAG	122
QY	45	LeuAlaThrLeuArgAlaGlnTyrProGlnValGluValArgCysGlyGluLeuAspAla	64
DB	123	CTGGCGCACCCCTGGTGGCCAGTATCCGAGGTGAAGTGGCTGGCGCGAATCTGCACGCC	182
QY	65	GluPheLeuCysSerAlaArgGluLeuTyrValSerProGlyLeuSerLeuArgThrPro	84
DB	183	GAGTTCCTCTGCTCGCGCCGCGAACTACTAGTCAAGCCCGCTTGTCGCTGCACACCCC	242
QY	85	AlaLeuValGlnAlaAlaAlaLysGlyValArgIleSerGlyAspIleAspLeuPheAla	104
DB	243	GGCTGGGTACAGGGCGCGCGAAAGGTGCGCATCTCCGGTGACATCGATCTTTCGGC	302
QY	105	ArgGluAlaLysAlaProIleValAlaIleThrGlySerAsnAlaLysSerThrVal	123
DB	303	CSCGAGGCGAAGGCCCGATCGTCCCATCACGGTTCCAAACGCCGAAGACACCGTG	359

## RESULT 6

US-09-252-991A-7624  
; Sequence 7624, Application US/09252991A  
; Patent No. 6551795

Alignment Scores:	2.22e-56	Length:	1359
Pred. No.:	67.00	Matches:	67
Score:	100.00%	Conservative:	0
Percent Similarity:	100.00%	Mismatches:	0
Best Local Similarity:	100.00%	Indels:	0
Query Match:	14.96%	Gaps:	0
DB:	4		

[US-09-701-229-2 (1-448) x [US-09-252-991A-7624 (1-1359)]

[illegible]

Db 62 GACGAAGACAGTCCGGCAGCGCCGAGCTGGCCCGGCGAAGCGGCGGTGTGTGCG 121  
 QY 422 ProAlaCysAlaSerLeuAspMetPheLysAsnPhelGluArgGlyAlaGLeuPheAla 441  
 Db 122 CCGCGCTGCGGAGCTGGACATGTTCAAGAACTTCGAAGACGGGACGCCCTGTGCGCC 181  
 QY 442 LysAlaValGluGluLeuAla 448  
 Db 182 AAAGCCGTAGAGGAGCTAGCG 202

RESULT 7  
 US-09-252-991A-7929/c  
 ; Sequence 729, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; FILE REFERENCE: 107196.136  
 ; CURRENT APPLICATION NUMBER: US/09/252,991A  
 ; PRIOR FILING DATE: 1999-02-18  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 7929  
 ; LENGTH: 1170  
 ; ORGANISM: Pseudomonas aeruginosa  
 US-09-252-991A-7929

Alignment Scores:  
 Pred No.: 2,92e-13 Length: 1170  
 Score: 23.00 Matches: 23  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 5.13% Indels: 0  
 Db: 4 Gaps: 0

US-09-701-229-2 (1-448) x US-09-252-991A-7929 (1-1170)  
 QY 1 MetSerLeuLeuAlaSerAspHisPheArgIleValValGlyLeuGlyLysSerglyMet 20  
 Db 71 ATGACGCTGATCGCTCCGACCACTTCGCGATCGTTGCGCTCGGCAAGAGCGGCATG 12  
 QY 21 SerLeuVal 23  
 Db 11 TCCCTGGTG 3

RESULT 8  
 US-09-328-352-1282  
 ; Sequence 1282, Application US/09328352  
 ; Patent No. 6562958  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Gary L. Breton et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
 ; FILE REFERENCE: GTC99-03PA  
 ; CURRENT APPLICATION NUMBER: US/09/328,352  
 ; CURRENT FILING DATE: 1999-06-04  
 ; NUMBER OF SEQ ID NOS: 8252  
 ; SEQ ID NO 1282  
 ; LENGTH: 1416  
 ; TYPE: DNA  
 ; ORGANISM: Acinetobacter baumannii  
 US-09-328-352-1282

Alignment Scores:  
 Pred No.: 2,7e-07 Length: 1416  
 Score: 17.00 Matches: 17  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0

Query Match: 3.79% Indels: 0  
 DB: .4 Gaps: 0  
 US-09-701-229-2 (1-448) x US-09-328-352-1282 (1-1416)  
 QY 110 ProIleValAlaIleThrGlySerAsnAlaLysSerThrValThrLeu 126  
 Db 385 CCGATTGCGGATTACAGTTCCAAATGCAAAAGTACTGTAACCACTTTA 435

RESULT 9  
 US-09-557-884-1  
 ; Sequence 1, Application US/09557884  
 ; Patent No. 6506581  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Fleischmann et al.  
 ; TITLE OF INVENTION: the Nucleotide sequence of the Haemophilus influenzae Rd Genome, Fragments Thereof, and Uses Thereof  
 ; NUMBER OF SEQUENCES: 1  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Human Genome Sciences, Inc.  
 ; STREET: 9410 Key West Avenue  
 ; CITY: Rockville  
 ; STATE: MD  
 ; COUNTRY: USA  
 ; ZIP: 20850  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: 3 1/2 inch diskette  
 ; COMPUTER: Dell Pentium  
 ; OPERATING SYSTEM: MS DOS v6.22  
 ; SOFTWARE: ASCII Text  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/557,884  
 ; FILING DATE: 25-Apr-2000  
 ; CLASSIFICATION: <Unknown>  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/476,102  
 ; FILING DATE: JUN-5-1995  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Michelle S. Marks  
 ; REGISTRATION NUMBER: 41,971  
 ; REFERENCE/DOCKET NUMBER: PBI86P3  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 301-309-8504  
 ; TELEFAX: 301-309-8439  
 ; INFORMATION FOR SEQ ID NO: 1:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1830121 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
 US-09-557-884-1

Alignment Scores:  
 Pred. No.: 0.212 Length: 1830121  
 Score: 14.00 Matches: 14  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 3.12% Indels: 0  
 DB: 4 Gaps: 0  
 US-09-701-229-2 (1-448) x US-09-557-884-1 (1-1830121)

QY 157 GluLeuTyrrValLeuGluLeuSerSerPheGluLeuThr 170  
 Db 1203588 GAACCTTATGCTAGAGCTTCTAGTTTTCAGCTTGAGACA 1203629  
 RESULT 10  
 US-09-643-990A-1  
 ; Sequence 1, Application US/09643990A  
 ; Patent No. 6528289  
 ; GENERAL INFORMATION:

```
; APPLICANT: Robert D. Fleischmann
; Mark D. Adams
; Owen White
; Hamilton O. Smith
; J. Craig Venter
;
; TITLE OF INVENTION: The Nucleotide sequence of
; the Haemophilus influenzae Rd Genome, Fragments
; Thereof, and Uses Thereof
;
; NUMBER OF SEQUENCES: 1
;
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville,
; STATE: MD
; COUNTRY: USA
; ZIP: 20850
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3 1/2 inch diskette
; COMPUTER: Dell Pentium
; OPERATING SYSTEM: MS DOS v6.22
; SOFTWARE: ASCII Text
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/643,990A
; FILING DATE: 23-Aug-2000
; CLASSIFICATION: <Unknown>
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/487,429
; FILING DATE: 1995-06-07
; APPLICATION NUMBER: 08/426,787
; FILING DATE: 1995-04-21
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Kenley K. Hoover
; REGISTRATION NUMBER: 40,302
; REFERENCE/DOCKET NUMBER: PB186PIC1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301-610-5790
; TELEFAX: 310-309-8439
;
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1830121 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
;
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-643-990A-1

Alignment Scores:
Pred. No.: 0.212 Length: 1830121
Score: 14.00 Matches: 14
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.12% Indels: 0
DB: 4 Gaps: 0

US-09-701-229-2 (1-448) x US-09-643-990A-1 (1-1830121)
157 GluLeuTyValLeuGluLeuSerPheGlnLeuGluThr.170
DB 1203588 GAACCTTATGTACTAGAGCTTCTTAGTTTACCTTGAGACA 1203629

RESULT 11
US-09-313-294A-6271/c
; Sequence 6271, Application US/09313294A
; Patent No. 6476212
; GENERAL INFORMATION:
; APPLICANT: Lalgudi, Raghunath V.
; APPLICANT: Ito, Laura Y.
; APPLICANT: Sherman, Bradley K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR
; FILE REFERENCE: PL-0017 US
; CURRENT APPLICATION NUMBER: US/09/313,294A
; CURRENT FILING DATE: 1999-05-14
; NUMBER OF SEQ ID NOS: 7600

; APPLICANT: Black, Michael
```

```
; SOFTWARE: PERL Program
; SEQ ID NO 6271
; LENGTH: 297
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. 6476212 700351520H1
; NAME/KEY: unsure
; LOCATION: 133-178
; OTHER INFORMATION: a, t, c, g, or other
US-09-313-294A-6271

Alignment Scores:
Pred. No.: 4.37 Length: 297
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.01% Indels: 0
DB: 4 Gaps: 0

US-09-701-229-2 (1-448) x US-09-313-294A-6271 (1-297)
QY 348 LeuValLeuAlaGlyAspGly 356
DB 261 CTCGTCCTCTCGCCGGTGGTGACGGC 235

RESULT 12
US-09-313-294A-5881/c
; Sequence 5881, Application US/09313294A
; Patent No. 6476212
; GENERAL INFORMATION:
; APPLICANT: Lalgudi, Raghunath V.
; APPLICANT: Ito, Laura Y.
; APPLICANT: Sherman, Bradley K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR
; FILE REFERENCE: PL-0017 US
; CURRENT APPLICATION NUMBER: US/09/313,294A
; CURRENT FILING DATE: 1999-05-14
; NUMBER OF SEQ ID NOS: 7600
; SOFTWARE: PERL Program
; SEQ ID NO 5881
; LENGTH: 298
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. 6476212 700350840H1
; NAME/KEY: unsure
; LOCATION: 58, 109-154, 263
; OTHER INFORMATION: a, t, c, g, or other
US-09-313-294A-5881

Alignment Scores:
Pred. No.: 4.38 Length: 298
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.01% Indels: 0
DB: 4 Gaps: 0

US-09-701-229-2 (1-448) x US-09-313-294A-5881 (1-298)
QY 348 LeuValLeuAlaGlyAspGly 356
DB 237 CTCGTCCTCTCGCCGGTGGTGACGGC 211

RESULT 13
US-08-936-165A-78
; Sequence 78, Application US/08936165A
; Patent No. 6348582
; GENERAL INFORMATION:
; APPLICANT: Black, Michael
```

APPLICANT: Burnham, Martin  
APPLICANT: Hodgson, John  
APPLICANT: Knowles, David  
APPLICANT: Lonetto, Michael  
APPLICANT: Nicholas, Richard  
APPLICANT: Pratt, Julie  
APPLICANT: Reichard, Richard  
APPLICANT: Rosenberg, Martin  
APPLICANT: Ward, Judith  
TITLE OF INVENTION: No. 6348582el Prokaryotic Polynucleotides,  
NUMBER OF INVENTION: Polypeptides and Their Uses  
CORRESPONDENCE ADDRESSES: 534  
ADDRESSEE: SmithKline Beecham Corporation  
STREET: 709 Swedeland Road  
CITY: King of Prussia  
STATE: PA  
COUNTRY: USA  
ZIP: 19406-0939  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/936,165A  
FILING DATE: 24-SEP-1997  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/027,032  
FILING DATE: 24-SEP-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Gimmi, Edward R  
REGISTRATION NUMBER: 38,891  
REFERENCE/DOCKET NUMBER: P50549  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 610-270-4478  
TELEFAX: 610-270-5090  
TELEX:  
INFORMATION FOR SEQ ID NO: 78:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 717 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: Genomic DNA  
US-08-936-165A-78

Alignment Scores:  
Pred. No.: 10.1 Length: 717  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00%  
Best Local Similarity: 100.00%  
Query Match: 2.01%  
Indels: 0  
DB: 4 Gaps: 0

US-09-701-229-2 (1-448) x US-08-936-165A-78 (1-717)

QY 418 ValLeuLeuSerProAlaCysAlaSer 426  
Db 411 GTATTATTGTCACCTGCTTGTGGAGT 437  
|||||

RESULT 14  
US-09-147-928-1  
Sequence 1, Application US/09147928  
Patent No. 6211161  
GENERAL INFORMATION:  
APPLICANT: Palmer, Leslie M.  
APPLICANT: Pratt, Julie M.  
APPLICANT: Hodgson, John E.  
APPLICANT: Beattie, David T.  
APPLICANT: Lowe, Adrian M.  
APPLICANT: Lonetto, Michael A.

APPLICANT: Nicholas, Richard O.  
APPLICANT: Deresiewicz, Robert L.  
TITLE OF INVENTION: MURD  
FILE REFERENCE: GM10106  
CURRENT APPLICATION NUMBER: US/09/147,928  
CURRENT FILING DATE: 1999-03-19  
EARLIER APPLICATION NUMBER: 60/061,064  
EARLIER FILING DATE: 1997-10-03  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 1  
LENGTH: 1242  
TYPE: DNA  
ORGANISM: Staphylococcus aureus  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)...(1239)  
NAME/KEY: misc\_feature  
LOCATION: (1)...(1242)  
OTHER INFORMATION: n = A,T,C or G  
US-09-147-928-1

Alignment Scores:  
Pred. No.: 17 Length: 1242  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00%  
Best Local Similarity: 100.00%  
Query Match: 2.01%  
Indels: 0  
DB: 3 Gaps: 0

US-09-701-229-2 (1-448) x US-09-147-928-1 (1-1242)

QY 418 ValLeuLeuSerProAlaCysAlaSer 426  
Db 1138 GTATTATTGTCACCTGCTTGTGGAGT 1164  
|||||

RESULT 15  
US-09-530-836-1  
Sequence 1, Application US/09530836  
Patent No. 6534284  
GENERAL INFORMATION:  
APPLICANT: Mohammed El-Sherbeini  
APPLICANT: Kenny Kin Wong  
APPLICANT: Wayne M. Geissler  
TITLE OF INVENTION: MURD PROTEIN AND GENE OF STAPHYLOCOCCUS  
TITLE OF INVENTION: AUREUS  
FILE REFERENCE: 19985P  
CURRENT APPLICATION NUMBER: US/09/530,836  
CURRENT FILING DATE: 2000-05-03  
PRIOR APPLICATION NUMBER: 60/064,183  
PRIOR FILING DATE: 1997-11-04  
PRIOR APPLICATION NUMBER: PCT/US98/23156  
PRIOR FILING DATE: 1998-10-30  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 1  
LENGTH: 1372  
TYPE: DNA  
ORGANISM: Staphylococcus aureus murd  
US-09-530-836-1

Alignment Scores:  
Pred. No.: 18.6 Length: 1372  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00%  
Best Local Similarity: 100.00%  
Query Match: 2.01%  
Indels: 0  
DB: 4 Gaps: 0

US-09-701-229-2 (1-448) x US-09-530-836-1 (1-1372)

QY 418 ValLeuLeuSerProAlaCysAlaSer 426  
|||||

Db 1258 GTATTATTGTCACCTGCTTGCGAGT 1284

Search completed: August 14, 2003, 11:33:45  
Job time : 1195 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 14, 2003, 09:38:07 ; Search time 350 Seconds

(without alignments)

2864.564 Million cell updates/sec

Title: us-09-701-229-2

Perfect score: 448

Sequence: 1 MSIIASHDFRIVVGLGKSGM.....MFKNFEEGRGLFAKAVEELA 448

Scoring table:

OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1504479 seqs, 1118970152 residues

Word size: 1

Total number of hits satisfying chosen parameters: 2626184

Minimum DB seq length: 25

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp  
-Q/cgn2\_1/USPTO.spool/US09701229/runat\_09082003\_161238\_5080/app.query.fasta\_1.647  
-DB=PublishedApplications\_NA -QFMT=fastap -SUFFIX=olip2n.rnpb -MINMATCH=0.1  
-LOOPCL=0 -LOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo  
-TRANS=human40.cdi -LIST=45 -DOALIGN=200 -THR\_SCORE=quality -THR\_MIN=1  
-ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=25  
-MAXLEN=2000000000 -USER=US09701229\_@CGN\_1\_1\_333\_@runat\_09082003\_161238\_5080  
-NCPU=6 -ICPU=3 -NO\_MMAP -LARGQUERY -NEG\_SCORES=0 -WAIT -DSBLOCK=100  
-LONGLOG -DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : PublishedApplications\_NA:

1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/1/pubpna/PCTUS\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq:\*  
10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq:\*  
12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*  
13: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq:\*  
14: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq:\*  
15: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*  
16: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*  
17: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Length	ID	Description
1	448	100.0	1347 9	US-09-815-242-7912 Sequence 7912, Ap

2	14	3.1	1314	9	US-09-815-242-7090	Sequence 7090, Ap
3	14	3.1	1317	9	US-09-741-669-259	Sequence 259, App
4	14	3.1	1317	9	US-09-815-242-5932	Sequence 5932, Ap
5	14	3.1	1317	9	US-09-815-242-9978	Sequence 9978, Ap
6	14	3.1	1830121	14	US-10-325-960-1	Sequence 1, Appli
c 7	9	2.0	135	9	US-09-815-242-3503	Sequence 3503, Ap
c 8	9	2.0	135	9	US-09-815-242-3597	Sequence 3597, Ap
c 9	9	2.0	216	9	US-09-815-242-1519	Sequence 1519, Ap
c 10	9	2.0	327	9	US-09-815-242-1487	Sequence 1487, Ap
c 11	9	2.0	383	9	US-09-815-242-2986	Sequence 2986, Ap
c 12	9	2.0	429	13	US-10-027-632-180203	Sequence 180203,
c 13	9	2.0	429	13	US-10-027-632-180204	Sequence 180204,
c 14	9	2.0	485	9	US-09-815-242-2037	Sequence 2037, Ap
c 15	9	2.0	498	13	US-10-027-632-280432	Sequence 280432,
c 16	9	2.0	710	13	US-10-027-632-14190	Sequence 14190, A
c 17	9	2.0	717	9	US-09-939-980-78	Sequence 78, Appl
c 18	9	2.0	719	13	US-10-027-632-16087	Sequence 16087, A
c 19	9	2.0	999	12	US-10-259-165-369	Sequence 369, App
c 20	9	2.0	1002	12	US-10-259-165-17	Sequence 17, Appl
c 21	9	2.0	1320	8	US-08-781-986A-80	Sequence 80, Appl
c 22	9	2.0	1347	9	US-09-815-242-4419	Sequence 4419, Ap
c 23	9	2.0	1350	9	US-09-815-242-8250	Sequence 8250, Ap
c 24	9	2.0	1350	9	US-09-815-242-8695	Sequence 8695, Ap
c 25	9	2.0	1491	14	US-10-156-761-3569	Sequence 3569, Ap
c 26	9	2.0	9025608	14	US-10-156-761-1	Sequence 1, Appli
c 27	8	1.8	25	14	US-10-098-263B-120644	Sequence 120644,
c 28	8	1.8	271	9	US-09-815-242-878	Sequence 878, App
c 29	8	1.8	282	14	US-10-156-761-5138	Sequence 5138, Ap
c 30	8	1.8	338	9	US-09-864-761-31989	Sequence 31989, A
c 31	8	1.8	370	13	US-10-027-632-68104	Sequence 68104, A
c 32	8	1.8	370	13	US-10-027-632-68105	Sequence 68105, A
c 33	8	1.8	388	10	US-09-960-352-41	Sequence 41, Appl
c 34	8	1.8	403	10	US-09-960-352-5964	Sequence 5964, Ap
c 35	8	1.8	403	10	US-09-960-352-11363	Sequence 11363, A
c 36	8	1.8	404	13	US-10-027-632-39955	Sequence 39955, A
c 37	8	1.8	404	13	US-10-027-632-39956	Sequence 39956, A
c 38	8	1.8	419	10	US-09-960-352-969	Sequence 969, App
c 39	8	1.8	467	11	US-09-918-995-21561	Sequence 21561, A
c 40	8	1.8	490	9	US-09-864-761-15475	Sequence 15475, A
c 41	8	1.8	519	13	US-10-027-632-85411	Sequence 85411, A
c 42	8	1.8	519	13	US-10-027-632-314877	Sequence 314877,
c 43	8	1.8	520	13	US-10-027-632-288812	Sequence 288812,
c 44	8	1.8	529	14	US-10-284-985-28	Sequence 28, Appl
c 45	8	1.8	541	9	US-09-864-761-14821	Sequence 14821, A

#### ALIGNMENTS

##### RESULT 1

US-09-815-242-7912  
Sequence 7912, Application US/09815242  
Patent No. US20020061569A1

##### GENERAL INFORMATION:

APPLICANT: Hasebeck, Robert  
APPLICANT: Ohlsen, Karl L.  
APPLICANT: Zyskind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
FILE REFERENCE: ELITRA.011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
CURRENT FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578

```

: PRIOR FILING DATE: 2000-10-23
: PRIOR APPLICATION NUMBER: 60/253,625
: PRIOR FILING DATE: 2000-11-27
: PRIOR APPLICATION NUMBER: 60/257,931
: PRIOR FILING DATE: 2000-12-22
: PRIOR APPLICATION NUMBER: 60/269,308
: PRIOR FILING DATE: 2001-02-16
: NUMBER OF SEQ ID NOS: 14110
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 7912
: LENGTH: 1347

```

Alignment Scores:

Pred. No.:	0	Length:	1347
Score:	448.00	Matches:	448
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	9	Gaps:	0
US-09-701-229-2 (1-448) x US-09-815-242-7912 (1-1347)			
QY	1	MetSerLeuIleAlaSerAspHisPheArgIleValValGlyLeuGlyLySerGlyMet	20
DB	1	ATGAGCCTGATCGCTCCGACCACTCCGCATCGTTGTCGGCCTCGCAAGAGCGCATG	60
QY	21	SerLeuValArgTyrLeuAlaArgGlyLeuProPheAlaValValAspThrArgGlu	400
DB	61	TCCCTGGTGGCTACTCGCGCGCGGGTTCCTTTCGCCGTGTCGATACCCGAGAG	120
QY	41	AsnProProGluLeuAlaThrLeuArgAlaGlnTyrProGlnValGluValArgCysGly	60
DB	121	AACCCCGCGAGCTGGCCACCTCGGTGCCAGTATCCGAGGTGAAGTGGCTTCGCG	180
QY	61	GluLeuAspAlaGluPheLeuCysSerAlaArgGluLeuTyrValSerProGlyLeuSer	80
DB	181	GAACTCGACGCGGAGTTCCTCTGCTCCGCCGCGAACTCTATGTCAGCCCCGGCTGTCG	240
QY	81	LeuArgThrProAlaLeuValGlnAlaAlaLysGlyValArgIleSerGlyAspIle	100
DB	241	CTGCGCACCCCTGGCGTGTACAGCGCGCGAAGGCGTGCATCTCCGGTGACATC	300
QY	101	AspLeuPheAlaArgGluAlaLysAlaProIleValAlaIleThrGlySerAsnAlaLys	120
DB	301	GATCTCTTTCGCGCGGAGCGAAGCCCGCATCGTCCCATCCCGTTCCAACGCGAAG	360
QY	121	SerThrValThrLeuValGlyGluMetAlaValAlaAlaAspLysArgValAlaVal	140
DB	361	AGCACCGTGACCCCTGGTGGCGCAATGGCGGTGCCCGCAAGCGGTGTCGCCGCT	420
QY	141	GlyGlyAsnLeuGlyThrProAlaLeuAspLeuLeuAlaAspAspIleGluLeuTyrVal	160
DB	421	GGCGCAACCTCGGCACCCCGCGCTCGACCTGCTGGCGACGACATCGAGCTGTACGTG	480
QY	161	LeuGluLeuSerSerPheGlnLeuGluThrCysAspArgLeuAsnAlaGluValAlaThr	180
DB	481	TTTGAGCTGTGAGCTTCCAGCTGGAACCTGGCGATCGCTCAACGCCGAGGTGGCGAC	540
QY	181	ValLeuAsnValSerGluAspHisMetAspArgTyrAspGlyMetAlaAspTyrHisLeu	200
DB	541	GTGCTGAACGTACGCGAAGACCAATGGATGGCTACGACGGCATGCTGACTACCACCTG	600
QY	201	AlaLysHisArgIlePheArgGlyAlaArgGlnValValValAsnArgAlaAspAlaLeu	220
DB	601	GCCAGCACCGGATCTCCCGGGTCCCGCAGGTCGTGGTGANTCGGCCGATGCCCTG	660
QY	221	ThrArgProLeuIleAlaAspThrValProCysTrpSerPheGlyLeuAsnLysProAsp	240

## RESULT 2

```

US-09-815-242-7090
; Sequence 7090, Application US/09815242
; Patent NO. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Ess
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23

```

;; PRIOR APPLICATION NUMBER: 60/253,625  
;; PRIOR FILING DATE: 2000-11-27  
;; PRIOR APPLICATION NUMBER: 60/257,931  
;; PRIOR FILING DATE: 2000-12-22  
;; PRIOR APPLICATION NUMBER: 60/269,308  
;; PRIOR FILING DATE: 2001-02-16  
;; NUMBER OF SEQ ID NOS: 14110  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 7090  
;; LENGTH: 1314  
;; TYPE: DNA  
;; ORGANISM: Haemophilus influenzae  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: (1)...(1314)  
US-09-815-242-7090

Alignment Scores:  
Pred. No.: 0.000202 Length: 1314  
Score: 14.00 Matches: 14  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 3.12% Indels: 0  
DB: 9 Gaps: 0

US-09-701-229-2 (1-448) x US-09-815-242-7090 (1-1314)

QY 157 GluLeuTyrrValLeuGluLeuSerPheGlnLeuGluThr 170  
|||||  
Db 457 GAACCTTATGACTAGAGCTTCTAGTTTCAGCTTGAGACA 498

## RESULT 3

US-09-741-669-259  
;; Sequence 259, Application US/09741669  
;; Patent No. US200200227181

;; GENERAL INFORMATION:  
;; APPLICANT: Forsyth, R. Allyn  
;; APPLICANT: Ohlsen, Kari L.  
;; APPLICANT: Zyskind, Judith W.  
;; TITLE OF INVENTION: Genes identified as required for  
;; proliferation of E. coli  
;; FILE REFERENCE: ELITRA.009A  
;; CURRENT APPLICATION NUMBER: US/09/741,669  
;; CURRENT FILING DATE: 2000-12-19  
;; PRIOR APPLICATION NUMBER: US 60/173005  
;; PRIOR FILING DATE: 1999-12-23  
;; NUMBER OF SEQ ID NOS: 481  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 259  
;; LENGTH: 1317  
;; TYPE: DNA  
;; ORGANISM: Escherichia coli  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: (1)...(1317)  
US-09-741-669-259

Alignment Scores:  
Pred. No.: 0.000203 Length: 1317  
Score: 14.00 Matches: 14  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 3.12% Indels: 0  
DB: 9 Gaps: 0

US-09-701-229-2 (1-448) x US-09-741-669-259 (1-1317)

QY 157 GluLeuTyrrValLeuGluLeuSerPheGlnLeuGluThr 170  
|||||  
Db 457 GAACCTGACGTGCTGGAAGTGTGAGCTCCAGCTGGAACC 498

## RESULT 4

US-09-815-242-5932

;; Sequence 5932, Application US/09815242  
;; Patent No. US20020061569A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Haselbeck, Robert  
;; APPLICANT: Ohlsen, Kari L.  
;; APPLICANT: Zyskind, Judith W.  
;; APPLICANT: Wall, Daniel  
;; APPLICANT: Trawick, John D.  
;; APPLICANT: Carr, Grant J.  
;; APPLICANT: Yamamoto, Robert T.  
;; APPLICANT: Xu, H. Howard  
;; TITLE OF INVENTION: Identification of Essential Genes in  
;; Prokaryotes  
;; FILE REFERENCE: ELITRA.011A  
;; CURRENT APPLICATION NUMBER: US/09/815,242  
;; CURRENT FILING DATE: 2001-03-21  
;; PRIOR APPLICATION NUMBER: 60/191,078  
;; PRIOR FILING DATE: 2000-03-21  
;; PRIOR APPLICATION NUMBER: 60/206,848  
;; PRIOR FILING DATE: 2000-05-23  
;; PRIOR APPLICATION NUMBER: 60/207,727  
;; PRIOR FILING DATE: 2000-05-26  
;; PRIOR APPLICATION NUMBER: 60/242,578  
;; PRIOR FILING DATE: 2000-10-23  
;; PRIOR APPLICATION NUMBER: 60/253,625  
;; PRIOR FILING DATE: 2000-11-27  
;; PRIOR APPLICATION NUMBER: 60/257,931  
;; PRIOR FILING DATE: 2000-12-22  
;; PRIOR APPLICATION NUMBER: 60/269,308  
;; PRIOR FILING DATE: 2001-02-16  
;; NUMBER OF SEQ ID NOS: 14110  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 5932  
;; LENGTH: 1317  
;; TYPE: DNA  
;; ORGANISM: Escherichia coli  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: (1)...(1317)  
US-09-815-242-5932

Alignment Scores:  
Pred. No.: 0.000203 Length: 1317  
Score: 14.00 Matches: 14  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 3.12% Indels: 0  
DB: 9 Gaps: 0

US-09-701-229-2 (1-448) x US-09-815-242-5932 (1-1317)

QY 157 GluLeuTyrrValLeuGluLeuSerPheGlnLeuGluThr 170  
|||||  
Db 457 GAACCTGACGTGCTGGAAGTGTGAGCTCCAGCTGGAACC 498

## RESULT 5

US-09-815-242-9978  
;; Sequence 9978, Application US/09815242  
;; Patent No. US20020061569A1

;; GENERAL INFORMATION:  
;; APPLICANT: Haselbeck, Robert  
;; APPLICANT: Ohlsen, Kari L.  
;; APPLICANT: Zyskind, Judith W.  
;; APPLICANT: Wall, Daniel  
;; APPLICANT: Trawick, John D.  
;; APPLICANT: Carr, Grant J.  
;; APPLICANT: Yamamoto, Robert T.  
;; APPLICANT: Xu, H. Howard  
;; TITLE OF INVENTION: Identification of Essential Genes in  
;; Prokaryotes  
;; FILE REFERENCE: ELITRA.011A  
;; CURRENT APPLICATION NUMBER: US/09/815,242  
;; CURRENT FILING DATE: 2001-03-21

```
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9978
; LENGTH: 1317
; TYPE: DNA
; ORGANISM: Salmonella typhi
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1317)
US-09-815-242-9978

Alignment Scores:
Pred. No.: 0.000203          Length: 1317
Score: 14.00                Matches: 14
Percent Similarity: 100.00%   Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.12%           Indels: 0
DB: 9                        Gaps: 0

US-09-701-229-2 (1-448) x US-09-815-242-9978 (1-1317)

QY 157 GluLeuTyrValLeuGluLeuSerPheGlnLeuGluThr 170
   |||||
Db 457 GRATTGCTGCTGGAATATCCAGTTCCATTGGAAACT 498

RESULT 6
US-10-329-960-1
; Sequence 1, Application US/10329960
; Publication No. US2003009277A1
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus influenzae Rd Genome, Frag
; TITLE OF INVENTION: Thereof, and Uses Thereof
; FILE REFERENCE: PB186P1
; CURRENT APPLICATION NUMBER: US/10/329,960
; CURRENT FILING DATE: 2003-01-02
; PRIOR APPLICATION NUMBER: US 09/643,990
; PRIOR FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: US 08/487,429
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/426,787
; PRIOR FILING DATE: 1995-04-21
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1830121
; TYPE: DNA
; ORGANISM: Haemophilus influenzae
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4747)..(4747)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (9921)..(9921)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (10150)..(10150)
; OTHER INFORMATION: n equals a, t, g or c

; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (29298)..(29298)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (36543)..(36543)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (36551)..(36551)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (36636)..(36636)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (40808)..(40810)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (44416)..(44416)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (44905)..(44905)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (44975)..(44975)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (45593)..(45593)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (45732)..(45732)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (47036)..(47036)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (51334)..(51334)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (51602)..(51602)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (51786)..(51786)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (51805)..(51805)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (55369)..(55369)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (65309)..(65309)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (65313)..(65313)
; OTHER INFORMATION: n equals a, t, g or c
```

FEATURE:	NAME/KEY: misc_feature	
LOCATION: (80024)..(80024)	OTHER INFORMATION: n equals a, t, g or c	
FEATURE:	NAME/KEY: misc_feature	
LOCATION: (100091)..(100091)	OTHER INFORMATION: n equals a, t, g or c	
FEATURE:	NAME/KEY: misc_feature	
LOCATION: (102696)..(102696)	OTHER INFORMATION: n equals a, t, g or c	
FEATURE:	NAME/KEY: misc_feature	
LOCATION: (105121)..(105121)	OTHER INFORMATION: n equals a, t, g or c	
FEATURE:	NAME/KEY: misc_feature	
LOCATION: (107248)..(107248)	OTHER INFORMATION: n equals a, t, g or c	
FEATURE:	NAME/KEY: misc_feature	
LOCATION: (117136)..(117136)	OTHER INFORMATION: n equals a, t, g or c	
FEATURE:	NAME/KEY: misc_feature	
LOCATION: (119750)..(119750)	OTHER INFORMATION: n equals a, t, g or c	
FEATURE:	NAME/KEY: misc_feature	
LOCATION: (119924)..(119924)	OTHER INFORMATION: n equals a, t, g or c	
FEATURE:	NAME/KEY: misc_feature	
LOCATION: (120038)..(120038)	OTHER INFORMATION: n equals a, t, g or c	
FEATURE:	NAME/KEY: misc_feature	
LOCATION: (121344)..(121344)	OTHER INFORMATION: n equals a, t, g or c	
FEATURE:	NAME/KEY: misc_feature	
LOCATION: (122167)..(122167)	OTHER INFORMATION: n equals a, t, g or c	
FEATURE:	NAME/KEY: misc_feature	
LOCATION: (122336)..(122336)	OTHER INFORMATION: n equals a, t, g or c	
FEATURE:	NAME/KEY: misc_feature	
LOCATION: (131340)..(131340)	OTHER INFORMATION: n equals a, t, g or c	
FEATURE:	NAME/KEY: misc_feature	
LOCATION: (131360)..(131360)	OTHER INFORMATION: n equals a, t, g or c	
FEATURE:	NAME/KEY: misc_feature	
LOCATION: (139910)..(139910)	OTHER INFORMATION: n equals a, t, g or c	
FEATURE:	NAME/KEY: misc_feature	
LOCATION: (140398)..(140398)	OTHER INFORMATION: n equals a, t, g or c	
FEATURE:	NAME/KEY: misc_feature	
LOCATION: (142750)..(142750)	OTHER INFORMATION: n equals a, t, g or c	
FEATURE:	NAME/KEY: misc_feature	
LOCATION: (145058)..(145058)	OTHER INFORMATION: n equals a, t, g or c	
FEATURE:	NAME/KEY: misc_feature	

```

; NAME/KEY: misc_feature
; LOCATION: (145171)..(145171)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (145942)..(145942)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (147197)..(147197)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (150841)..(150841)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (152500)..(152500)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (152530)..(152530)

Alignment Scores:
Pred. No.: 0.129 Length: 1830121
Score: 14.00 Matches: 14
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.12% Indels: 0
DB: 14 Gaps: 0

US-09-701-229-2 (1-448) x US-10-329-960-1 (1-1830121)

Qy 157 GluLeuYrYallLeuGluLeuSerPheClnLeuGluThr 170
Db 1203358 GAACTTATGTACTAGAGCTTCTAGTTTCAGCTTGAGACA 1203629

RESULT 7
US-09-815-242-3503/c
; Sequence 3503, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3503
; LENGTH: 135

```

;  
PRIOR FILING DATE: 2000-03-21



;; PRIOR APPLICATION NUMBER: 60/206,848  
;; PRIOR FILING DATE: 2000-05-23  
;; PRIOR APPLICATION NUMBER: 60/207,727  
;; PRIOR FILING DATE: 2000-05-26  
;; PRIOR APPLICATION NUMBER: 60/242,578  
;; PRIOR FILING DATE: 2000-10-23  
;; PRIOR APPLICATION NUMBER: 60/253,625  
;; PRIOR FILING DATE: 2000-11-27  
;; PRIOR APPLICATION NUMBER: 60/257,931  
;; PRIOR FILING DATE: 2000-12-22  
;; PRIOR APPLICATION NUMBER: 60/269,308  
;; PRIOR FILING DATE: 2001-02-16  
;; NUMBER OF SEQ ID NOS: 14110  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 1487  
;; LENGTH: 327  
;; TYPE: DNA  
;; ORGANISM: Staphylococcus aureus  
US-09-815-242-1487

Alignment Scores:  
Pred. No.: 6.85 Length: 327  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.01% Indels: 0  
DB: 9 Gaps: 0

US-09-701-229-2 (1-448) x US-09-815-242-1487 (1-327)

QY 418 ValLeuLeuSerProAlaCysAlaSer 426  
|||||  
DB 145 GTATTATTGTCACCTGCTGTGCGAGT 119

## RESULT 11

;; Sequence 2986, Application US/09815242  
;; Patent No. US20020061569A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Haselbeck, Robert  
;; APPLICANT: Ohlsen, Kari L.  
;; APPLICANT: Zyskind, Judith W.  
;; APPLICANT: Wall, Daniel  
;; APPLICANT: Trawick, John D.  
;; APPLICANT: Carr, Grant J.  
;; APPLICANT: Yamamoto, Robert T.  
;; APPLICANT: Xu, H. Howard  
;; TITLE OF INVENTION: Identification of Essential Genes in  
;; FILE REFERENCE: ELITRA.011A  
;; CURRENT APPLICATION NUMBER: US/09/815,242  
;; CURRENT FILING DATE: 2001-03-21  
;; PRIOR APPLICATION NUMBER: 60/191,078  
;; PRIOR FILING DATE: 2000-03-21  
;; PRIOR APPLICATION NUMBER: 60/206,848  
;; PRIOR FILING DATE: 2000-05-23  
;; PRIOR APPLICATION NUMBER: 60/207,727  
;; PRIOR FILING DATE: 2000-05-26  
;; PRIOR APPLICATION NUMBER: 60/242,578  
;; PRIOR FILING DATE: 2000-10-23  
;; PRIOR APPLICATION NUMBER: 60/253,625  
;; PRIOR FILING DATE: 2000-11-27  
;; PRIOR APPLICATION NUMBER: 60/257,931  
;; PRIOR FILING DATE: 2000-12-22  
;; PRIOR APPLICATION NUMBER: 60/269,308  
;; PRIOR FILING DATE: 2001-02-16  
;; NUMBER OF SEQ ID NOS: 14110  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 2986  
;; LENGTH: 383  
;; TYPE: DNA  
;; ORGANISM: Staphylococcus aureus  
US-09-815-242-2986

Alignment Scores:  
Pred. No.: 7.89 Length: 383  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.01% Indels: 0  
DB: 9 Gaps: 0

US-09-701-229-2 (1-448) x US-09-815-242-2986 (1-383)

QY 418 ValLeuLeuSerProAlaCysAlaSer 426  
|||||  
DB 39 GTATTATTGTCACCTGCTGTGCGAGT 13

## RESULT 12

;; Sequence 180203, Application US/10027632  
;; GENERAL INFORMATION:  
;; APPLICANT: Wang, David G.  
;; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
;; FILE REFERENCE: 108827.129  
;; CURRENT APPLICATION NUMBER: US/10/027,632  
;; CURRENT FILING DATE: 2002-04-30  
;; PRIOR APPLICATION NUMBER: US 60/218,006  
;; PRIOR FILING DATE: 2000-07-12  
;; PRIOR APPLICATION NUMBER: US 60/198,676  
;; PRIOR FILING DATE: 2000-04-20  
;; PRIOR APPLICATION NUMBER: US 60/193,483  
;; PRIOR FILING DATE: 2000-03-29  
;; PRIOR APPLICATION NUMBER: US 60/185,218  
;; PRIOR FILING DATE: 2000-02-24  
;; PRIOR APPLICATION NUMBER: US 60/167,363  
;; PRIOR FILING DATE: 1999-11-23  
;; PRIOR APPLICATION NUMBER: US 60/156,358  
;; PRIOR FILING DATE: 1999-09-28  
;; PRIOR APPLICATION NUMBER: US 60/146,002  
;; PRIOR FILING DATE: 1999-08-09  
;; NUMBER OF SEQ ID NOS: 325720  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 180203  
;; LENGTH: 429  
;; TYPE: DNA  
;; ORGANISM: Human  
US-10-027-632-180203

Alignment Scores:  
Pred. No.: 8.73 Length: 429  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.01% Indels: 0  
DB: 13 Gaps: 0

US-09-701-229-2 (1-448) x US-10-027-632-180203 (1-429)

QY 312 TrpValArgGluArgGlnGlyValSer 320  
|||||  
DB 208 TGGGTGAGAGAAAGACAGGTTGTGAGT 182

## RESULT 13

;; Sequence 180204, Application US/10027632  
;; GENERAL INFORMATION:  
;; APPLICANT: Wang, David G.  
;; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
;; FILE REFERENCE: 108827.129  
;; CURRENT APPLICATION NUMBER: US/10/027,632  
;; CURRENT FILING DATE: 2002-04-30  
;; PRIOR APPLICATION NUMBER: US 60/218,006  
;; PRIOR FILING DATE: 2000-07-12

;; PRIOR APPLICATION NUMBER: US 60/198,676  
;; PRIOR FILING DATE: 2000-04-20  
;; PRIOR APPLICATION NUMBER: US 60/193,483  
;; PRIOR FILING DATE: 2000-03-29  
;; PRIOR APPLICATION NUMBER: US 60/185,218  
;; PRIOR FILING DATE: 2000-02-24  
;; PRIOR APPLICATION NUMBER: US 60/167,363  
;; PRIOR FILING DATE: 1999-11-23  
;; PRIOR APPLICATION NUMBER: US 60/156,358  
;; PRIOR FILING DATE: 1999-09-28  
;; PRIOR APPLICATION NUMBER: US 60/146,002  
;; PRIOR FILING DATE: 1999-08-09  
;; NUMBER OF SEQ ID NOS: 325720  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 180204  
;; LENGTH: 429  
;; TYPE: DNA  
;; ORGANISM: Human  
US-10-027-632-180204

Alignment Scores:  
Pred. No.: 8.73 Length: 429  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservatives: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.01% Indels: 0  
DB: 13 Gaps: 0

US-09-701-229-2 (1-448) x US-10-027-632-180204 (1-429)

QY 312 TrpValArgGluArgGlnGlyValSer 320  
DB 208 TGGGTGAGAGAAGACAGCGGTGTGAGT 182

## RESULT 14

US-09-815-242-2037/c  
;; Sequence 2037, Application US/09815242  
;; Patent No. US20020061569A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Haselbeck, Robert  
;; APPLICANT: Ohlsen, Kari L.  
;; APPLICANT: Zyskind, Judith W.  
;; APPLICANT: Wall, Daniel  
;; APPLICANT: Trawick, John D.  
;; APPLICANT: Carr, Grant J.  
;; APPLICANT: Yamamoto, Robert T.  
;; APPLICANT: Xu, H. Howard  
;; TITLE OF INVENTION: Identification of Essential Genes in  
;; FILE REFERENCE: ELITRA.011A  
;; CURRENT APPLICATION NUMBER: US/09/815,242  
;; CURRENT FILING DATE: 2001-03-21  
;; PRIOR APPLICATION NUMBER: 60/191,078  
;; PRIOR FILING DATE: 2000-03-21  
;; PRIOR APPLICATION NUMBER: 60/206,848  
;; PRIOR FILING DATE: 2000-05-23  
;; PRIOR APPLICATION NUMBER: 60/207,727  
;; PRIOR FILING DATE: 2000-05-26  
;; PRIOR APPLICATION NUMBER: 60/242,578  
;; PRIOR FILING DATE: 2000-10-23  
;; PRIOR APPLICATION NUMBER: 60/253,625  
;; PRIOR FILING DATE: 2000-11-27  
;; PRIOR APPLICATION NUMBER: 60/257,931  
;; PRIOR FILING DATE: 2000-12-22  
;; PRIOR APPLICATION NUMBER: 60/269,308  
;; PRIOR FILING DATE: 2001-02-16  
;; NUMBER OF SEQ ID NOS: 14110  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 2037  
;; LENGTH: 485  
;; TYPE: DNA  
;; ORGANISM: Staphylococcus aureus  
US-09-815-242-2037

Alignment Scores:  
Pred. No.: 9.74 Length: 485  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservatives: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.01% Indels: 0  
DB: 9 Gaps: 0

US-09-701-229-2 (1-448) x US-09-815-242-2037 (1-485)

QY 418 ValLeuLeuSerProAlaCysAlaSer 426  
DB 287 GTATTATTGTACCTGCTGTGCGAGT 261

## RESULT 15

US-10-027-632-280432  
;; Sequence 280432, Application US/10027632  
;; GENERAL INFORMATION:  
;; APPLICANT: Wang, David G.  
;; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
;; FILE REFERENCE: 108827.129  
;; CURRENT APPLICATION NUMBER: US/10/027,632  
;; CURRENT FILING DATE: 2002-04-30  
;; PRIOR APPLICATION NUMBER: US 60/218,006  
;; PRIOR FILING DATE: 2000-07-12  
;; PRIOR APPLICATION NUMBER: US 60/198,676  
;; PRIOR FILING DATE: 2000-04-20  
;; PRIOR APPLICATION NUMBER: US 60/193,483  
;; PRIOR FILING DATE: 2000-03-29  
;; PRIOR APPLICATION NUMBER: US 60/185,218  
;; PRIOR FILING DATE: 2000-02-24  
;; PRIOR APPLICATION NUMBER: US 60/167,363  
;; PRIOR FILING DATE: 1999-11-23  
;; PRIOR APPLICATION NUMBER: US 60/156,358  
;; PRIOR FILING DATE: 1999-09-28  
;; PRIOR APPLICATION NUMBER: US 60/146,002  
;; PRIOR FILING DATE: 1999-08-09  
;; NUMBER OF SEQ ID NOS: 325720  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 280432  
;; LENGTH: 498  
;; TYPE: DNA  
;; ORGANISM: Human  
US-10-027-632-280432

Alignment Scores:  
Pred. No.: 9.98 Length: 498  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservatives: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.01% Indels: 0  
DB: 13 Gaps: 0

US-09-701-229-2 (1-448) x US-10-027-632-280432 (1-498)

QY 224 LeuIleAlaAspThrValProCysTrp 232  
DB 347 TTGATCGCTGATCTGTTCCATGCTGG 373

Search completed: August 14, 2003, 11:30:57  
Job time : 918 secs

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 14, 2003, 09:10:41 ; Search time 2291 Seconds  
(without alignments)  
4752.687 Million cell updates/sec

Title: US-09-701-229-2  
Perfect score: 448  
Sequence: 1 MSLIASDHFRIVVGLGKSGM.....MFKNFERGRLFAKAVEELA 448  
Scoring table: OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0  
Searched: 22781392 seqs, 12152238056 residues  
Word size: 1  
Total number of hits satisfying chosen parameters: 45546778  
Minimum DB seq length: 25  
Maximum DB seq length: 200000000  
Post-processing: Listing first 45 summaries  
Command line parameters:  
-MODEL=frame+p2n.model -DRV=xlp  
-Q/cgn2\_1/USPTO.spool/US09701229/runat\_09082003\_161237\_5033/app\_query.fasta\_1.647  
-DB=EST -Qfmt=fastap -SUFFIX=olip2n.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45  
-DOALIGN=200 -THR\_SCORE=quality -THR\_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto  
-NORM=ext -HEAPSIZ=500 -MINLEN=25 -MAXLEN=200000000  
-USER=US09701229.ecgn\_1.1.4237 @runat\_09082003\_161237\_5033 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :  
EST :  
1: em.estha :  
2: em.esthum :  
3: em.estin :  
4: em.estnu :  
5: em.estov :  
6: em.estpl :  
7: em.estro :  
8: em.htc :  
9: gb.estl :  
10: gb.est2 :  
11: gb.htc :  
12: gb.est3 :  
13: gb.est4 :  
14: gb.est5 :  
15: em.estfun :  
16: em.estom :  
17: em.gss\_hum :  
18: em.gss\_inv :  
19: em.gss\_pln :  
20: em.gss\_vrt :  
21: em.gss\_fun :  
22: em.gss\_man :  
23: em.gss\_mus :  
24: em.gss\_pro :  
25: em.gss\_rod :  
26: em.gss\_phg :  
27: em.gss\_vri :  
28: gb.gasl :

29: gb.gss2 :\*

pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
C 1	200	44.6	1445	29	BZ568959	BZ568959 pacs2-164
C 2	186	41.5	892	29	BZ579116	BZ579116 msh2_6205
C 3	41	9.2	1171	29	BZ559806	BZ559806 pacs2-164
C 4	21	4.7	452	10	BE184674	BE184674 PM4-HT068
C 5	11	2.5	928	29	BZ567557	BZ567557 pacs2-164
C 6	9	2.0	199	6	AL814293	AL814293 Triticum
C 7	9	2.0	279	12	BM370877	BM370877 EBRO04_SQ
C 8	9	2.0	309	14	CB605003	CB605003 3529_1_64
C 9	9	2.0	349	14	D39320	D39320 RICR3327A R
C 10	9	2.0	359	14	C19538	C19538 C19538 Rice
C 11	9	2.0	360	10	BB870897	BB870897 BB870897
C 12	9	2.0	366	14	CA622878	CA622878 wlin.pk01
C 13	9	2.0	366	14	CA692865	CA692865 wlm96.pk0
C 14	9	2.0	380	9	AI544865	AI544865 fb69b08.x
C 15	9	2.0	380	28	AZ060728	AZ060728 RPCI-23-4
C 16	9	2.0	392	10	BE403471	BE403471 WHE0429.H
C 17	9	2.0	392	12	BI781117	BI781117 EBma03_SQ
C 18	9	2.0	411	12	BI777969	BI777969 EBRO07_SQ
C 19	9	2.0	424	9	AA510452	AA510452 VQ32f06.f
C 20	9	2.0	429	9	AU162673	AU162673 AU162673
C 21	9	2.0	431	9	AW315553	AW315553 12985 MAR
C 22	9	2.0	442	9	AU183975	AU183975 AU183975
C 23	9	2.0	453	9	AW019231	AW019231 f461a10.x
C 24	9	2.0	456	9	AU064507	AU064507 AU064507
C 25	9	2.0	472	10	AI667240	AI667240 fd1a109.x
C 26	9	2.0	477	10	BG605535	BG605535 WHE2239.B
C 27	9	2.0	484	10	BF200711	BF200711 WHE0821-0
C 28	9	2.0	487	12	BG907641	BG907641 TaLr1161F
C 29	9	2.0	497	9	AW564083	AW564083 LG1_282.C
C 30	9	2.0	498	12	BM377980	BM377980 EBma04_SQ
C 31	9	2.0	499	9	AA136741	AA136741 zK99a05.f
C 32	9	2.0	512	28	AZ098624	AZ098624 RPCI-23-2
C 33	9	2.0	515	12	BM323975	BM323975 PTCL_29.H
C 34	9	2.0	517	10	BE640499	BE640499 945004C03
C 35	9	2.0	531	28	AZ769198	AZ769198 LM0569M03
C 36	9	2.0	533	14	CB878729	CB878729 HP09316T
C 37	9	2.0	538	12	BM534621	BM534621 fx69e10.Y
C 38	9	2.0	540	14	CA700810	CA700810 wkmlc.pk0
C 39	9	2.0	549	12	BM859675	BM859675 fy59f07.x
C 40	9	2.0	559	10	BE510292	BE510292 946051F02
C 41	9	2.0	562	14	CA618126	CA618126 wlin.pk00
C 42	9	2.0	562	14	CB426323	CB426323 601498.WA
C 43	9	2.0	566	10	BG605315	BG605315 WHE2331.C
C 44	9	2.0	569	14	CA686112	CA686112 wlm96.pk0
C 45	9	2.0	579	13	BU371697	BU371697 603597031

ALIGNMENTS

RESULT 1  
BZ568959/c  
LOCUS  
DEFINITION BZ568959 1445 bp DNA linear GSS 17-DEC-2002  
pac2-164\_8171.y2 pacs2-164 Pseudomonas aeruginosa genomic clone  
BZ568959  
ACCESSION BZ568959  
VERSION BZ568959.1 GI:27202799  
KEYWORDS GSS.  
SOURCE Pseudomonas aeruginosa  
ORGANISM Pseudomonas aeruginosa  
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
Pseudomonadaceae; Pseudomonas.  
REFERENCE 1 (bases 1 to 1445)

AUTHORS Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M., Burns, J.L., Kaul, R. and Olsen, M.V.  
 TITLE Whole-Genome-Sequence variation among multiple isolates of Pseudomonas aeruginosa library  
 JOURNAL J. Bacteriol., (2002) In press  
 COMMENT Contact: Chris K. Raymond  
 Genome Center  
 University of Washington  
 Box 352145, Seattle, WA 98105-2145, USA  
 Tel: 2062216954  
 Fax: 2066857244  
 Email: craymond@u.washington.edu  
 Class: shotgun.

FEATURES source  
 1. 1445  
 /organism="Pseudomonas aeruginosa"  
 /mol\_type="genomic DNA"  
 /strain="2-164"  
 /db\_xref="taxon:287"  
 /clone\_lib="pacs2-164"  
 /note="clinical isolate 2-164 Whole genomic shotgun library."

BASE COUNT 308 a 424 c 412 g 293 t 8 others  
 ORIGIN

Alignment Scores:  
 Pred. No.: 2.63e-179 Length: 1445  
 Score: 200.00 Matches: 200  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 44.64% Indels: 0  
 DB: 29 Gaps: 0

US-09-701-229-2 (1-448) x B2568959 (1-1445)

QY 147 ProAlaLeuAspLeuAlaAspAspLeuGluLeuTyrValLeuGluLeuSerSerPhe 166  
 DB 675 CCGCGCTCGACCTCTCTGGCGACGACATCTGAGCTGTACGTGTGGAGCTTCGAGCTTC 616  
 QY 167 GlnLeuGluThrCysAspArgLeuAsnAlaGluValAlaThrValLeuAsnValSerGlu 186  
 DB 615 CAGCTGGAAACCTGGATCGCTCAACCGCGAGGTGGCGACCGTGTCAAGCTCAGCGAA 556  
 QY 187 AspHisMetAspArgTyrAspGlyMetAlaAspTyrHisLeuAlaLysHisArgIlePhe 206  
 DB 555 GACCATATGGATCGCTACGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 496  
 QY 207 ArgGlyAlaArgGlnValValAlaValAlaValAlaLeuThrArgProLeuIleAla 226  
 DB 495 CGCGGTGCGCGCGAGTCTGT 436  
 QY 227 AspThrValProCysTrpSerPheGlyLeuAsnLysProAspPheLysAlaPheGlyLeu 246  
 DB 435 GATACCGTGGCTGT 376  
 QY 247 IleGluGluAspGlyGlnLysTrpLeuAlaPheGlnPheAspLysLeuLeuProValGly 266  
 DB 375 ATCAGAGGAAGACGGCCAGAACTGGCTGGCGTTCAGTTTCGACAAAGCTGCTCGCGTTGGC 316  
 QY 267 GluLeuLysIleArgGlyAlaHisAsnTyrSerAsnAlaLeuAlaAlaLeuAlaLeuGly 286  
 DB 315 GAACCTGAACATCGGTGGCGCCACCACTATTTCACACGCGCTCGCGCGCTGGCGTGGGC 256  
 QY 287 HisAlaValGlyLeuProPheAspAlaMetLeuGlyAlaLeuLysAlaPheSerGlyLeu 306  
 DB 255 CATCGGTGCGCTGCTGTGCGCCATGCTCGCGCGCTGAGAGCGTTTCGCGCGCTG 196  
 QY 307 AlaHisArgCysGlnTrpValArgGluArgGlnGlyValSerTyrTyrAspAspSerLys 326  
 DB 195 GCTCATCGCTCCAGTGGGTACGGAGCGGCGGCGGTGAGCTACTACGACGATTTCGAAG 136  
 QY 327 AlaThrAsnValGlyAlaAlaLeuAlaAlaIleGluGlyLeuGlyAlaAspIleAspGly 346

DB 135 GCCACCAACGTCGCGCGCTCTGGCGGATCGAGGGCTGGTGTCCGACATCGACGGG 76  
 RESULT 2  
 B2579116/c  
 LOCUS msh2\_6205.y2 msh Pseudomonas aeruginosa genomic clone msh2\_6205,  
 DEFINITION genomic survey sequence.  
 ACCESSION B2579116  
 VERSION B2579116.1 GI:27214177  
 KEYWORDS GSS.  
 SOURCE Pseudomonas aeruginosa  
 ORGANISM Pseudomonas aeruginosa  
 Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
 Pseudomonadaceae; Pseudomonas.  
 REFERENCE 1 (bases 1 to 892)  
 AUTHORS Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M., Burns, J.L., Kaul, R. and Olsen, M.V.  
 TITLE Whole-Genome-Sequence variation among multiple isolates of Pseudomonas aeruginosa library  
 JOURNAL J. Bacteriol., (2002) In press  
 COMMENT Contact: Chris K. Raymond  
 Genome Center  
 University of Washington  
 Box 352145, Seattle, WA 98105-2145, USA  
 Tel: 2062216954  
 Fax: 2066857244  
 Email: craymond@u.washington.edu  
 Class: shotgun.

FEATURES source  
 1. 892  
 /organism="Pseudomonas aeruginosa"  
 /mol\_type="genomic DNA"  
 /strain="MSH"  
 /db\_xref="taxon:287"  
 /clone\_lib="msh2\_6205"  
 /note="Environmental isolate. Whole genomic shotgun library."

BASE COUNT 173 a 277 c 276 g 160 t 6 others  
 ORIGIN

Alignment Scores:  
 Pred. No.: 3.16e-166 Length: 892  
 Score: 186.00 Matches: 186  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 41.52% Indels: 0  
 DB: 29 Gaps: 0

US-09-701-229-2 (1-448) x B2579116 (1-892)

QY 91 AlaLysGlyValArgIleSerGlyAspIleAspLeuPheAlaArgGluAlaLysAlaPro 110  
 DB 642 CGCAAAGCGTGGCATCTCGGTGACATCATCTTCGCCCGGAGCGAGCGCCG 583  
 QY 111 IleValAlaIleThrGlySerAsnAlaLysSerThrValThrLeuValGlyClyMet 130  
 DB 582 ATCTGTCGCATCACCGGTTCACGCGAAGACACCGTGTGACCCCTGGTGGCGAAATG 523  
 QY 131 AlaValAlaAlaAspLysArgValAlaValGlyGlyAsnLeuGlyThrProAlaLeuAsp 150  
 DB 522 CGGTGGCGCGGCAACAGCGTGTGCGCGTGGCGGCAACCTCGGCGCGCTCGAC 463  
 QY 151 LeuLeuAlaAspAspIleGluLeuTyrValLeuGluLeuSerSerPheGlnLeuGluThr 170  
 DB 462 CTGCTGGCGGACGACATCGAGCTGTACGTGTGTGGAGCTGTGAGCTTCCAGCTGGAACC 403  
 QY 171 CysAspArgLeuAsnAlaGluValAlaThrValLeuAsnValSerGluAspHisMetasp 190  
 DB 402 TGCATCGCTCAACGCGGAGGTGGCGACCGCTGTGAACGTCAGGAGACCATATGGAT 343  
 QY 191 ArgTyrAspGlyMetAlaAspTyrHisLeuAlaLysHisArgIlePheArgGlyAlaArg 210

```

Db      342 CGCTACAGCGCATGGCTGACTACCACTGCGCAAGCACCAGGATCTTCGCGGTGCCGC -283
QY      211 GlnValValAlaAsnArgAlaAspAlaLeuThrArgProLeuLeuAlaAspThrValPro 230
Db      282 CAGGTGCTGTGAATCGCGCGATGCCCTGACCCGACCGCTGATCGCCGATACCGTCCG 223
QY      231 CysTrpSerPheGlyLeuAsnLysProAspPheLysAlaPheGlyLeuLeuGluAsp 250
Db      222 TGTGTGCTGTCGGCTGACACAGCGGACTTCAAGGCTTCGCGCTGATCGAGGAGAC 163
QY      251 GlyGlnLysTrpLeuAlaPheGlnPheAspLysLeuLeuProValGlyGluLeuLysIle 270
Db      162 GGCACAGAGTGGTGGCTTCCAGTTCGACACAGCTGTCGCGGTGGCGAAGTCAAGATC 103
QY      271 ArgGlyAlaHisAsnTyr 276
Db      102 CGTGGCGGCCACCAACTAT 85

RESULT 3
LOCUS      BZ559806
DEFINITION pacs2-164_1863.x1 pacs2-164 Pseudomonas aeruginosa genomic clone
ACCESSION BZ559806
VERSION    BZ559806.1 GI:27176698
KEYWORDS   GSS.
SOURCE     Pseudomonas aeruginosa
ORGANISM   Pseudomonas aeruginosa
REFERENCE  1 (bases 1 to 1171)
AUTHORS    Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
            Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE      Whole-Genome-Sequence variation among multiple isolates of
            Pseudomonas aeruginosa library
JOURNAL    J. Bacteriol., (2002) In press
COMMENT     Contact: Chris K. Raymond
            Genome Center
            University of Washington
            Box 352145, Seattle, WA 98105-2145, USA
            Tel: 20622216954
            Fax: 2066857244
            Email: craymond@u.washington.edu
            Class: shotgun.
FEATURES   source
            1..1171
            /organism="Pseudomonas aeruginosa"
            /mol_type="genomic DNA"
            /strain="2-164"
            /db_xref="taxon:287"
            /clone="pacs2-164_1863"
            /clone_lib="pacs2-164"
            /note="clinical isolate 2-164 Whole genomic shotgun
            library."
BASE COUNT 218 a 325 c 329 g 298 t 1 others
ORIGIN
Alignment Scores:
Pred. No.: 3.47e-28 Length: 1171
Score: 41.00 Matches: 68
Percent Similarity: 97.14% Conservative: 0
Best Local Similarity: 97.14% Mismatches: 1
Query Match: 9.15% Indels: 2
DB: 29 Gaps: 0

US-09-701-229-2 (1-448) x BZ559806 (1-1171)
QY      310 CysGlnTrpValArgGluArgGlnGlyValSerTyrTyrAspAspSerLysAlaThrAsn 329
Db      150 TGCAGTGGGTAGCGGAGCGGAGCGGTGAGTACTACGACGATTCACAGGCCACCAAC 209
QY      330 ValGlyAlaAlaLeuAlaAlaIleGluGlyLeuGlyAlaAspIleAspGlyLysLeuVal 349

```

```

Db      210 GTCCGGCCGCCCTGGCGGCGATCGAGGGGTGGTCCGACATCGACGCGCAAGCTGGTG 269
QY      350 LeuLeu-AlaGlyGlyAspGlyLysGlyAlaAspPheHisAspLeuArgGluProValAl 369
Db      270 CTGAT-CGCGCGGCGAGACGCGCAAGGCGCGGATTTTCATGACTCGCGAGCGGTCGC 328
QY      369 aATGPhcCysArgAlaValValLeuLeu 378
Db      329 GCCTTCTGCGGCGGCTGTACTGCTT 356

RESULT 4
LOCUS      BE184674
DEFINITION PM4-HT0688-050500-002-d10 HT0688 Homo sapiens cDNA, mRNA sequence.
ACCESSION BE184674
VERSION    BE184674.1 GI:8663858
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
REFERENCE  1 (bases 1 to 452)
AUTHORS    Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
TITLE      Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
JOURNAL    Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE    20202663
PUBMED     10737800
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=et2-PM4-HT0688-050
            500-002-d10&t3=2000-05-05&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 26
            High quality sequence stop: 89.
FEATURES   Location/Qualifiers
            1..452
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /dev_stage="Adult"
            /clone_lib="HT0688"
            /note="Organ: head,neck; Vector: puc18; Site:1: SmaI;
            Site:2: SmaI; A mini-library was made by cloning products
            derived from ORESTES PCR (U.S. Letters patent application
            No. 196,716 - Ludwig Institute for Cancer Research)
            profiles into the pUC 18 vector. Reverse transcription of
            tissue mRNA and cDNA amplification were performed under
            low stringency conditions."
BASE COUNT 91 a 140 c 133 g 88 t
ORIGIN
Alignment Scores:
Pred. No.: 1.25e-09 Length: 452
Score: 21.00 Matches: 21
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.69% Indels: 0
DB: 10 Gaps: 0

```

US-09-701-229-2 (1-448) x BE184674 (1-452)

QY 277 SerAsnAlaLeuAlaLeuAlaLeuGlyHisAlaValGlyLeuPropheAspAlaMet 296  
 Db 241 TCCATGCGCTGGCGGCTTGGCACTCGGTACGCGGTGGCTGCCATTCGACCCCATG 300

QY 297 Leu 297

Db 301 CTC 303

# RESULT 5

BZ567557/c

LOCUS BZ567557 928 bp DNA linear GSS 17-DEC-2002  
 DEFINITION pacs2-164\_7094.y2 pacs2-164 Pseudomonas aeruginosa genomic clone  
 pacs2-164\_7094, genomic survey sequence.

ACCESSION BZ567557

VERSION BZ567557.1 GI:27199005

KEYWORDS GSS.

SOURCE Pseudomonas aeruginosa

## ORGANISM

Pseudomonas aeruginosa

Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;

Pseudomonadaceae; Pseudomonas.

1 (bases 1 to 928)

Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,

Burns, J.L., Kaul, R. and Olsen, M.V.

Whole-Genome-Sequence variation among multiple isolates of

Pseudomonas aeruginosa library

J. Bacteriol., (2002) In press

Contact: Chris K. Raymond

Genome Center

University of Washington

Box 352145, Seattle, WA 98105-2145, USA

Tel: 2062216954

Fax: 2066857244

Email: craymond@u.washington.edu

Class: shotgun.

FEATURES

source

1..928

/organism="Pseudomonas aeruginosa"

/mol\_type="genomic DNA"

/strain="2-164"

/db\_xref="taxon:287"

/clone="pacs2-164\_7094"

/clone\_lib="pacs2-164"

/note="clinical isolate 2-164 Whole genomic shotgun library."

BASE COUNT 238 a 228 c 236 g 226 t

## ORIGIN

Alignment Scores:

Pred. No.: 9.12 Length: 928

Score: 11.00 Matches: 11

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 2.46% Indels: 0

DB: 29 Gaps: 0

US-09-701-229-2 (1-448) x BZ567557 (1-928)

QY 279 AlaLeuAlaAlaLeuAlaLeuGlyHisAlaVal 289

Db 324 GCGCTTGGCGCTGGCGGTGGCGCATCGGTC 292

# RESULT 6

AL814293/c

ID AL814293 standard; RNA; EST; 199 BP.

XX

AC AL814293;

XX

SV AL814293.1

XX

DT 05-JUL-2002 (Rel. 72, Created)

DT 02-SEP-2002 (Rel. 72, Last updated, Version 2)

XX Triticum aestivum (mercia) clone E12\_e512\_plate\_7 maternal tissue (minus endosperm and embryo) 8 days post anthesis

XX EST; expressed sequence tag.

XX

OS Triticum aestivum (bread wheat)

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;

XX Triticaceae; Triticum.

XX

RN

RA Wilson I., Beswick R., Shepherd S., Barker G., Parker J., Owen P.,

RA Edwards D., Coghill J., Holdsworth M., Lenton J., Shewry P., Edwards K.,

RT "A BBSRC-funded wheat EST resource for the academic community";

XX unpublished.

XX

RN

RP 1-199

RA Barker G.;

RT

RL Submitted (03-JUL-2002) to the EMBL/GenBank/DBJ databases.

RL Barker G., Institute of Arable Crop Research, Long Ashton, Bristol BS41 9AF

XX United Kingdom.

XX

PH Key

Location/Qualifiers

1..199

/db\_xref="taxon:4565"

/mol\_type="mRNA"

/organism="Triticum aestivum"

/cultivar="mercia"

/dev\_stage="8 days post anthesis"

/tissue\_type="maternal tissue (minus endosperm and embryo)"

/clone\_lib="e:512"

/clone="E12\_e512\_plate\_7"

XX

SQ Sequence 199 BP; 48 A; 56 C; 73 G; 22 T; 0 other;

Alignment Scores:

Pred. No.: 129 Length: 199

Score: 9.00 Matches: 9

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 2.01% Indels: 0

DB: 6 Gaps: 0

US-09-701-229-2 (1-448) x AL814293 (1-199)

QY 348 LeuValLeuAlaGlyAspGly 356

Db 88 CTCGTCTCTCTCGCGGTGGTGACGGC 62

RESULT 7

BM370877/c

LOCUS BM370877.2

DEFINITION

Hordeum vulgare subsp. vulgare

sequence.

BM370877

VERSION

BM370877.2

KEYWORDS

EST.

SOURCE

ORGANISM

Hordeum vulgare subsp. vulgare

Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae

; Triticeae; Hordeum.

REFERENCE

1 (bases 1 to 279)

AUTHORS

Hedley, P., Liu, H., Caldwell, D., McCallum, N., Mudie, S., Cardie, L.,

Ramsay, L., Machray, G., Marshall, D.F.M. and Waugh, R.

Development of Barley Transcriptome Resources

Unpublished



COMMENT On Jan 10, 2002 this sequence version replaced gi:18114267.  
 Contact: Waugh R, Marshall DF  
 Genome Dynamics/Computational Biology  
 Scottish Crop Research Institute  
 Invergowrie, Dundee, DD2 5DA, Scotland, UK  
 Tel: 00 44 1382 562731  
 Fax: 00 44 1382 562426  
 Email: est@scri.sari.ac.uk  
 All sequence has a Phred quality score of 20 or over  
 Seq primer: M13 reverse.  
 Location/Qualifiers  
 1. .279  
 /organism="Hordeum vulgare subsp. vulgare"  
 /mol\_type="mRNA"  
 /cultivar="Optic"  
 /db\_xref="taxon:112509"  
 /clone="EBro04\_SQ002\_K07"  
 /tissue\_type="root"  
 /dev\_stage="3 week"  
 /lab\_host="PH108"  
 /notes="lib="root, 3 week, salt-stressed, cv Optic, EBro04"  
 /note="Vector: pSPORT1; Site.1: Sal I; Site.2: Not I;  
 Non-normalised library, directionally cloned into pSPORT1.  
 Derived from roots of 3 week old salt stressed barley  
 plants. Developed as part of the barley transcriptome  
 resources of BBSC/SEERAD funded cereal IGF (investigating  
 Gene Function) project."

BASE COUNT 48 a 103 c 77 g 51 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 189 Length: 279  
 Score: 9.00 Matches: 9  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.01% Indels: 0  
 DB: 12 Gaps: 0

US-09-701-229-2 (1-448) x BM370877 (1-279)  
 QY 397 ArgValAlaThrLeuAspGluAlaVal 405  
 |||||  
 Db 66 CGCGTCGACACCTTGACGAAGCAGC 40

RESULT 8  
 CB605003/c  
 LOCUS  
 DEFINITION CB605003 309 bp mRNA linear EST 16-MAY-2003  
 labs Zea mays cDNA, mRNA sequence.  
 CB605003  
 VERSION  
 KEYWORDS EST.  
 SOURCE  
 ORGANISM  
 Zea mays  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
 clade; Panicoideae; Andropogoneae; Zea.  
 1 (bases 1 to 309)  
 Walbot, V.  
 Maize ESTs from various cDNA libraries sequenced at Stanford  
 University  
 Unpublished  
 Contact: Walbot V  
 Department of Biological Sciences  
 Stanford University  
 855 California Ave, Palo Alto, CA 94304, USA  
 Tel: 650 723 2227  
 Fax: 650 725 8221  
 Email: walbot@stanford.edu  
 Plate: 3529.1\_64.1 row: E column: 07.  
 Location/Qualifiers  
 1. .309  
 /organism="Zea mays"

/mol\_type="mRNA"  
 /cultivar="B73"  
 /db\_xref="taxon:4577"  
 /tissue\_type="ear"  
 /dev\_stage="2 mm"  
 /lab\_host="E. coli XL0LR"  
 /clone\_lib="3529 - 2 mm ear tissue from Schmidt and Hake  
 labs"  
 /note="Organ: ear; Vector: PAD-GAL4-2.1; Site.1: EcoRI;  
 Site.2: XhoI; RNA isolated by Hake lab. 1 million pfu  
 amplified. Ampicillin is the selection marker."  
 BASE COUNT 72 a 88 c 113 g 36 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 212 Length: 309  
 Score: 9.00 Matches: 9  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.01% Indels: 0  
 DB: 14 Gaps: 0

US-09-701-229-2 (1-448) x CB605003 (1-309)  
 QY 348 LeuValLeuAlaGlyAspGly 356  
 |||||  
 Db 137 CTCGCTCTCTCGCGGTGTGACGCG 111

RESULT 9  
 D39320/c  
 LOCUS  
 DEFINITION D39320 349 bp mRNA linear EST 03-APR-2002  
 RICR3327A Rice root Oryza sativa (Japonica cultivar-group) cDNA,  
 mRNA sequence.  
 D39320  
 VERSION  
 KEYWORDS EST.  
 SOURCE  
 ORGANISM  
 Oryza sativa (Japonica cultivar-group)  
 Oryza sativa (Japonica cultivar-group)  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Ehrhartoideae; Oryzaceae; Oryza.  
 1 (bases 1 to 349)  
 Sasaki, T., Miyao, A. and Yamamoto, K.  
 Rice cDNA from callus 1995  
 Unpublished  
 Contact: Takuji Sasaki  
 National Institute of Agrobiological Resources  
 Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki  
 305-8602, Japan  
 Tel: 81-298-38-7441  
 Fax: 81-298-38-7468  
 Email: tsasaki@abrr.affrc.go.jp, URL: http://rgp.dna.affrc.go.jp/  
 PROJECT "RGP", Location/Qualifiers  
 1. .349  
 /organism="Oryza sativa (Japonica cultivar-group)"  
 /mol\_type="mRNA"  
 /cultivar="Nipponbare"  
 /db\_xref="taxon:39947"  
 /clone\_lib="Rice root"  
 /note="Prepared from seedling root."  
 BASE COUNT 72 a 110 c 118 g 44 t 5 others  
 ORIGIN

Alignment Scores:  
 Pred. No.: 243 Length: 349  
 Score: 9.00 Matches: 9  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.01% Indels: 0  
 DB: 14 Gaps: 0

US-09-701-229-2 (1-448) x D39320 (1-349)

QY 348 LeuValLeuAlaGlyGlyAspGly 356  
 |||||||  
 Db 228 CTCGCTCCTCTCGCGGTGGTACGCGC 202

RESULT 10  
 C19538/c  
 LOCUS  
 DEFINITION C19538 Rice panicle at ripening stage Oryza sativa (japonica cultivar-group) cDNA clone E10575\_1A, mRNA sequence.

ACCESSION C19538  
 VERSION C19538.1 GI:1631809  
 KEYWORDS EST.  
 SOURCE Oryza sativa (japonica cultivar-group)  
 ORGANISM Oryza sativa (japonica cultivar-group)  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.  
 REFERENCE 1 (bases 1 to 359)  
 AUTHORS Sasaki,T.  
 TITLE Rice cDNA from panicle at ripening stage  
 JOURNAL Unpublished  
 COMMENT Contact: Takuji Sasaki  
 National Institute of Agrobiological Resources  
 Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan  
 Tel: 81-298-38-7441  
 Fax: 81-298-38-7468  
 Email: tsasaki@abr.affrc.go.jp, URL: http://rgp.dna.affrc.go.jp/  
 PROJECT "RGP"

FEATURES  
 source  
 1..359  
 /organism="Oryza sativa (japonica cultivar-group)"  
 /mol\_type="mRNA"  
 /cultivar="Nipponbare"  
 /db\_xref="taxon:39947"  
 /clone="E10575\_1A"  
 /dev\_stage="ripening stage"  
 /clone\_lib="Rice panicle at ripening stage"  
 /note="organ: panicle; Rice cDNA from panicle at ripening stage"  
 BASE COUNT 72 a 107 c 119 g 50 t 11 others  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 251 Length: 359  
 Score: 9.00 Matches: 9  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.01% Indels: 0  
 DB: 14 Gaps: 0

US-09-701-229-2 (1-448) x C19538 (1-359)

QY 348 LeuValLeuAlaGlyGlyAspGly 356  
 |||||||  
 Db 174 CTCGCTCCTCTCGCGGTGGTACGCGC 148

RESULT 11  
 BB870897/c  
 LOCUS  
 DEFINITION BB870897 RIKEN full-length enriched, 10 days pregnant adult female ovary and uterus Mus musculus cDNA clone G630025M20 5', mRNA sequence.

ACCESSION BB870897  
 VERSION BB870897.1 GI:17117107  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 360)  
 AUTHORS Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,

Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.  
 RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al. 2001)

# TITLE JOURNAL COMMENT

Unpublished  
 Contact: Yoshihide Hayashizaki  
 Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute  
 The Institute of Physical and Chemical Research (RIKEN)  
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
 Tel: 81-45-503-9222  
 Fax: 81-45-503-9216  
 Email: genome-res@gsc.riken.go.jp,  
 URL: http://genome.gsc.riken.go.jp/  
 M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
 Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)  
 wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.  
 RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)  
 Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.  
 Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
 Please visit our web site (http://genome.gsc.riken.go.jp) for further details.  
 e mouse tissues.

# FEATURES source

1..360  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="G630025M20"  
 /sex="female"  
 /tissue\_type="ovary and uterus"  
 /dev\_stage="10 days pregnant adult"  
 /clone\_lib="RIKEN full-length enriched, 10 days pregnant adult female ovary and uterus"  
 BASE COUNT 96 a 112 c 85 g 67 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 252 Length: 360  
 Score: 9.00 Matches: 9  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.01% Indels: 0  
 DB: 10 Gaps: 0

US-09-701-229-2 (1-448) x BB870897 (1-360)

QY 348 LeuValLeuAlaGlyGlyAspGly 356  
 |||||||

Db 52 TTGGTCTTCGCTGGTGGCGACGGC 26

# RESULT 12 CA622878/c LOCUS

DEFINITION CA622878  
 wln.p.k0102.f2 wln Triticum aestivum cDNA clone wln.p.k0102.f2 5', end, mRNA sequence.

ACCESSION CA622878  
 VERSION CA622878.1 GI:25201174  
 KEYWORDS EST  
 SOURCE Triticum aestivum (bread wheat)  
 ORGANISM Triticum aestivum

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae ; Triticeae; Triticum.  
 1 (bases 1 to 366)  
 TINGEY,S.V., POWELL,W., WOLTERS,P., DOLAN,M., HAINEY,C., YUAN,Z., MAO,G., CARAHAN,N. and HANAFAY,M.K.  
 AUTHORS DuPont Wheat cDNA Sequence  
 TITLE DuPont Wheat cDNA Sequence  
 JOURNAL Unpublished  
 COMMENT Contact: Scott V. Tingey  
 Crop Genetics  
 E. I. DuPont de Nemours and Company  
 1 Innovation Way, P.O. Box 6104, Newark, DE 19714-6104, USA  
 Tel: 302-631-2602  
 Fax: 302-631-2607  
 Email: Scott.V.Tingey@USA.dupont.com  
 Seq primer: M13.

FEATURES  
 source Location/Qualifiers  
 1..366  
 /organism="Triticum aestivum"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:4565"  
 /clone="wlin.pk0102.f2"  
 /tissue\_type="leaf"  
 /clone\_lib="wlin"  
 /note="vector: pBluescript SK+; Site\_1: EcoRI; Site\_2: XhoI; Wheat (Triticum aestivum L.) leaf 7 day old seedling  
 XhoI; light grown (normalized)"  
 BASE COUNT 93 a 94 c 114 g 63 t 2 others  
 ORIGIN

Alignment Scores:  
 Pred. No.: 256 Length: 366  
 Score: 9.00 Matches: 9  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.01% Indels: 0  
 DB: 14 Gaps: 0

US-09-701-229-2 (1-448) x CA622878 (1-366)  
 QY 348 LeuValLeuLeuAlaGlyGlyAspGly 356  
 |||||  
 Db 79 CTCGCTCTCTCTCGCGGGTGGACGCC 53

RESULT 13  
 CA622665/c 366 bp mRNA linear EST 25-NOV-2002  
 LOCUS wlm96.pk060.b5 wlm96 Triticum aestivum cDNA clone wlm96.pk060.b5 5'  
 DEFINITION end, mRNA sequence.  
 ACCESSION CA622665  
 VERSION CA622665.1 GI:2541444  
 KEYWORDS EST  
 SOURCE Triticum aestivum (bread wheat)  
 ORGANISM Triticum aestivum

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae ; Triticeae; Triticum.  
 1 (bases 1 to 366)  
 TINGEY,S.V., POWELL,W., WOLTERS,P., DOLAN,M., HAINEY,C., YUAN,Z., MAO,G., CARAHAN,N. and HANAFAY,M.K.  
 AUTHORS DuPont Wheat cDNA Sequence  
 TITLE DuPont Wheat cDNA Sequence  
 JOURNAL Unpublished  
 COMMENT Contact: Scott V. Tingey  
 Crop Genetics  
 E. I. DuPont de Nemours and Company  
 1 Innovation Way, P.O. Box 6104, Newark, DE 19714-6104, USA  
 Tel: 302-631-2602  
 Fax: 302-631-2607

Email: Scott.V.Tingey@USA.dupont.com  
 Seq primer: M13.  
 Location/Qualifiers  
 1..366  
 /organism="Triticum aestivum"  
 /mol\_type="mRNA"  
 /cultivar="Stephens"  
 /db\_xref="taxon:4565"  
 /clone="wlm96.pk060.b5"  
 /tissue\_type="leaf"  
 /clone\_lib="wlm96"  
 /note="vector: pBluescript SK+; Site\_1: EcoRI; Site\_2: XhoI; Wheat (Triticum aestivum L.) seedlings 96 hr after inoculation w/ E. graminis"  
 BASE COUNT 65 a 125 c 114 g 53 t 9 others  
 ORIGIN

Alignment Scores:  
 Pred. No.: 256 Length: 366  
 Score: 9.00 Matches: 9  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.01% Indels: 0  
 DB: 14 Gaps: 0

US-09-701-229-2 (1-448) x CA692665 (1-366)  
 QY 348 LeuValLeuLeuAlaGlyGlyAspGly 356  
 |||||  
 Db 266 CTCGCTCTCTCTCGCGGGTGGACGCC 240

RESULT 14  
 AI544865/c 380 bp mRNA linear EST 07-JUN-2001  
 LOCUS fb69b08.xl zebrafish WashU MPIMG EST Danio rerio cDNA clone  
 DEFINITION IMAGE:3717111 3', mRNA sequence.  
 ACCESSION AI544865  
 VERSION AI544865.1 GI:4462238  
 KEYWORDS EST  
 SOURCE Danio rerio (zebrafish)  
 ORGANISM Danio rerio  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes ; Cyprinidae; Danio.  
 REFERENCE 1 (bases 1 to 380)  
 AUTHORS Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy ,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood ,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.  
 TITLE WashU zebrafish EST Project 1998  
 JOURNAL Unpublished  
 COMMENT Contact: Stephen L. Johnson  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: zbrafish@watson.wustl.edu  
 cDNA Library Preparation: Matthew Clark. cDNA Library Arrayed by: Matthew Clark. DNA Sequencing by: Washington University Genome Sequencing Center. Clone distribution: Genome Systems, St. Louis, Missouri (web address: www.genomesystems.com) (email contact: info@genomesystems.com) and Research Genetics, Huntsville, Alabama (web address: www.resgen.com) (email contact: info@resgen.com) and RessourcenzentrumPrimarDatenbank, Berlin, Germany (web address: www.rzpd.de)  
 Seq primer: T7 ET from Amersham  
 High quality sequence stop: 333  
 POLYA-No. Location/Qualifiers  
 1..380  
 /organism="Danio rerio"

/mol\_type="mrna"  
/db\_xref="taxon:7955"  
/clone\_image:3717111  
/sex="mixed"  
/tissue\_type="26 somite embryos, adult livers, shield  
stage embryos"  
/lab\_host="XLI-blue MRF"  
/clone\_lib="zebrafish WashU MPIMG EST"  
/note="Vector: pSPORT1; Site\_1: NotI; Site\_2: SalI; 1st  
strand cDNA was primed with a Not I - oligo(GT)15 primer  
[5'pGACTGTCCTAGATCGGAGCGGCCCTTTTCTTTT3'];  
double-stranded cDNA was ligated to Sal I adaptors (BRL),  
digested with Not I and cloned into the Not I and Sal I  
sites of the pSPORT1 vector (BRL). Library was constructed  
by Matthew Clark (Lehrach lab; ICRF, London and Max Planck  
Institut fuer Molekulare Genetik, Berlin). cDNAs for EST  
analysis were selected following oligonucleotide  
hybridization fingerprinting of arrayed clones from  
zebrafish late somitogenesis (26 ss), adult liver or  
embryonic shield stage (5.6 h) libraries. Fingerprint  
data were used to computationally cluster cDNAs, and a  
single cDNA from each cluster was chosen for sequencing.  
In some cases multiple members of the same cluster were  
sequenced to assess clustering parameters or single clones  
were sequenced additional times to assess quality  
control."

BASE COUNT 119 a 78 c 82 g 101 t  
ORIGIN

Alignment Scores:  
Pred. No.: 267 Length: 380  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.01% Indels: 0  
DB: 9 Gaps: 0

US-09-701-229-2 (1-448) x A1544865 (1-380)

QY 78 GlyLeuSerLeuArgThProAlaLeu 86

DB 106 GCAGTGGATTTCAGAACCCCTGCCCTA 80

RESULT 15

AZ060728/c

LOCUS RPCI-23-415J24.TV RPCI-23 Mus musculus genomic clone RPCI-23-415J24 380 bp DNA linear GSS 30-MAR-2000

DEFINITION , genomic survey sequence.

ACCESSION AZ060728

VERSION AZ060728.1 GI:7351899

KEYWORDS GSS.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

1 (bases 1 to 380)

Zhao, S., Mierman, W., Feldblyum, T., Malek, J., Shatsman, S., Akinret

, B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P.

and Fraser, C.M.

Mouse BAC End Sequences from Library RPCI-23

Other GSSs: RPCI-23-415J24.TJ

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhaoetigr.org

Clones are derived from the mouse BAC library RPCI-23. For BAC

library availability, please contact Pieter de Jong

(pieteredejong.med.buffalo.edu). Clones may be purchased from

BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)

or from Resea ch Genetics (info@resgen.com). BAC end page:  
http://www.tigr.org/tdb/bac\_ends/mouse/bac\_end\_intro.html  
Plate: 415 row: J column: 24

Seq primer: T7

Class: BAC ends.

FEATURES Location/Qualifiers

source

1..380

/organism="Mus musculus"

/mol\_type="genomic DNA"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="RPCI-23-415J24"

/sex="Female"

/lab\_host="DH10B"

/clone\_lib="RPCI-23"

/note="Organ: Kidney/Brain; Vector: pBACE3.6; Site\_1:  
EcoRI; Site\_2: EcoRI; Female C57BL/6J mouse kidney and/or  
brain genomic DNA was isolated and partially digested  
with a combination of EcoRI and EcoRI Methylase. Size  
selected DNA was cloned into the pBACE3.6 vector at the  
EcoRI sites. The ligation products were transformed into  
DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT 132 a 71 c 88 g 89 t

ORIGIN

Alignment Scores:

Pred. No.: 267 Length: 380  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.01% Indels: 0  
DB: 28 Gaps: 0

US-09-701-229-2 (1-448) x AZ060728 (1-380)

QY 277 SerAsnAlaLeuAlaAlaLeuAlaLeu 285

DB 328 TCAAAATGCTTTAGCAGCTTGGCATTG 302

Search completed: August 14, 2003, 11:13:47

Job time : 2305 secs

and is derived by analysis of the total score distribution.

SUMMARIES

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 14, 2003, 07:48:07 ; Search time 350 Seconds  
(without alignments)  
3455.281 Million cell updates/sec

Title: US-09-701-229-2  
Perfect score: 448  
Sequence: 1 MSLIASDHFRIVVGLKSGM.....MFKNFERGRLFAKAVEELA 448

Scoring table: OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2552756 seqs, 1349719017 residues  
Word size: 1

Total number of hits satisfying chosen parameters: 3293711

Minimum DB seq length: 25  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:  
-MODEL=frame+p2n.model -dbv=xlp  
-Q=Ctrl2\_1/USPTO\_spool/US09701229/runat\_09082003\_161235\_5011/app\_query.fasta.1.647  
-DB=N\_Geneseq\_19Jun03 -Qfmt=fastp -SUFFIX=olip2n.rng -MINMATCH=0.1 -LOOPEL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=oligo -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR\_SCORE=quality -THR\_MIN=1 -ALIGN=15 -MODE=LOCAL  
-OUTFTW=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=25 -MAXLEN=2000000000  
-USER=US09701229.ecgn.1.1.471.eurnat\_09082003\_161235\_5011 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : N\_Geneseq\_19Jun03:.\*  
1: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:.\*  
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:.\*  
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:.\*  
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:.\*  
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:.\*  
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:.\*  
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:.\*  
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:.\*  
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:.\*  
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:.\*  
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:.\*  
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:.\*  
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:.\*  
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:.\*  
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:.\*  
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:.\*  
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:.\*  
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:.\*  
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:.\*  
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:.\*  
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:.\*  
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:.\*  
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:.\*  
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:.\*  
25: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2003.DAT:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed,

Result No.	Score	Query Match	Length	DB ID	Description
1	448	100.0	1347	23	AAS54275 Pseudomonas aerugi
2	448	100.0	1450	21	AAZ47131 Pseudomonas aerugi
3	19	4.2	96109	22	AAZ28548 Genomic fragment #
4	14	3.1	1080	23	AAS89258 DNA encoding novel
5	14	3.1	1080	23	AAS89801 DNA encoding novel
6	14	3.1	1080	23	AAS90166 DNA encoding novel
7	14	3.1	1314	22	AAS3453 Haemophilus influe
8	14	3.1	1317	22	AAH81460 Escherichia coli p
9	14	3.1	1317	23	AAS52295 E. coli DNA for ce
10	14	3.1	1317	23	AAS56341 Salmonella typhi D
11	14	3.1	2058	23	AAS73012 DNA encoding novel
12	14	3.1	3117	23	AAS60995 DNA encoding novel
13	14	3.1	3117	23	AAS89240 DNA encoding novel
14	14	3.1	4831	23	AAS94542 DNA encoding novel
15	14	3.1	4942	23	AAS89816 DNA encoding novel
16	14	3.1	4944	23	AAS89273 DNA encoding novel
17	14	3.1	4944	23	AAS90181 DNA encoding novel
18	14	3.1	1830121	17	AAT42063 Haemophilus influe
19	11	2.5	2703	23	AAS64579 DNA encoding novel
20	11	2.5	2703	23	AAS66661 DNA encoding novel
21	10	2.2	356	21	AAZ53173 Neisseria meningit
22	10	2.2	356	21	AAZ53174 Neisseria meningit
23	10	2.2	1335	25	ABZ41528 N. gonorrhoeae nuc
24	10	2.2	1380	21	AAZ53175 Neisseria meningit
25	10	2.2	69936	21	AAA81479 N. meningitidis pa
26	10	2.2	349980	21	AAF21607 Neisseria meningit
27	10	2.2	1437668	21	AAA81490 N. meningitidis B
28	9	2.0	135	23	AAS50926 Staphylococcus aur
29	9	2.0	135	23	AAS51020 Staphylococcus aur
30	9	2.0	216	23	AAS48942 Staphylococcus aur
31	9	2.0	285	21	AAZ53157 Neisseria gonorrh
32	9	2.0	285	21	AAZ53158 Neisseria meningit
33	9	2.0	285	21	AAZ53159 Neisseria meningit
34	9	2.0	297	25	ABX87811 Corn ear-derived p
35	9	2.0	298	25	ABX87421 Corn ear-derived p
36	9	2.0	327	23	AAS48910 Staphylococcus aur
37	9	2.0	383	23	AAS50409 Staphylococcus aur
38	9	2.0	485	23	AAS49460 Staphylococcus aur
39	9	2.0	717	18	AAZ83927 DNA encoding a udp
40	9	2.0	717	19	AAV53378 DNA encoding a udp
41	9	2.0	1188	22	AAH52631 S. epidermidis ope
42	9	2.0	1242	20	AAX28684 Nucleotide sequenc
43	9	2.0	1320	18	AAV74391 Staphylococcus aur
44	9	2.0	1347	23	AAS51837 Staphylococcus aur
45	9	2.0	1350	23	AAS54613 Staphylococcus aur

ALIGNMENTS

RESULT 1  
AAS54275  
ID AAS54275 standard; DNA; 1347 BP.  
AC AAS54275;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE Pseudomonas aeruginosa DNA for cellular proliferation protein #406.  
XX  
KW Antisense; ds; prokaryotic cellular proliferation gene;  
KW antibiotic; antibacterial; drug design.  
XX  
OS Pseudomonas aeruginosa.  
XX  
FN WO200170955-A2.  
XX

PD 27-SEP-2001:  
 XX 21-MAR-2001; 2001WO-US09180.  
 XX 21-MAR-2000; 2000US-191078P.  
 PR 23-MAY-2000; 2000US-206848P.  
 PR 26-MAY-2000; 2000US-207272P.  
 PR 23-OCT-2000; 2000US-242578P.  
 PR 27-NOV-2000; 2000US-253625P.  
 PR 22-DEC-2000; 2000US-257931P.  
 PR 16-FEB-2001; 2001US-269308P.  
 XX (ELIT-) ELITRA PHARM INC.  
 XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;  
 PI Yamamoto RT, Xu HH;  
 XX WPI; 2001-611495/70.  
 DR P-PSDB; ANU36416.  
 XX  
 PT New polynucleotides for the identification and development of  
 PT antibiotics, comprise sequences of antisense nucleic acids -  
 XX  
 PS Claim 27; Seq ID No 7912; 511pp; English.  
 CC The invention relates to antisense inhibitors of genes essential to  
 CC prokaryotic cellular proliferation, their use in identifying the  
 CC genes, their use in the discovery of novel antibiotics, the essential  
 CC genes themselves and the encoded proteins. The prokaryotes used are  
 CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella  
 CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The  
 CC invention is also useful for the identification of potential new targets  
 CC for antibiotic development. The antisense nucleic acids can also be used  
 CC to identify proteins used in proliferation, to express these proteins,  
 CC and to obtain antibodies capable of binding to the expressed proteins.  
 CC The proteins can be used to screen compounds in rational drug discovery  
 CC programmes. The antisense nucleic acid sequence is also useful to screen  
 CC for homologous nucleic acids which are required for cell proliferation in  
 CC a wide variety of organisms. The present sequence encodes an  
 CC essential prokaryotic cellular proliferation protein.  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 1347 BP; 214 A; 453 C; 453 G; 227 T; 0 other;

Alignment Scores:  
 Pred. No.: 0 Length: 1347  
 Score: 448.00 Matches: 448  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 23 Gaps: 0

US-09-701-229-2 (1-448) x AAS54275 (1-1347)

Qy 1 MetSerLeuLeuAlaSerAspHisPheArgIleValValGlyLeuGlyLysSerGlyMet 20  
 Db 1 ATGAGCCTGATCGCTCCGACCACTTCGGCATCGTTGTTCGGCCTCGGCAAGCGGCATG 60  
 Qy 21 SerLeuValArgTyrLeuAlaArgGlyLeuProPheAlaValAspThrArgGlu 40  
 Db 61 TCCCTGGTGGCTACCTCGGCGCGCGGCTTTCGCGCTTCGCGATACCCGAGAG 120  
 Qy 41 AsnProProGluLeuAlaThrLeuArgAlaGlnTyrProGlnValGluValArgCysGly 60  
 Db 121 AACCCGGCGAGCTGGCCACCTCGTGGCCAGTATCCGAGGTGGAGTGGTTCGGCG 180  
 Qy 61 GluLeuAspAlaGluPheLeuCysSerAlaArgGluLeuTyrValSerProGlyLeuSer 80  
 Db 181 GAACCTGACGCGAGTTCCTCTGCTCCGCGCGCAACTCTATGTCAGCCCGGCTTCTCG 240

Qy 81 LeuArgThrProAlaLeuValGlnAlaAlaLysGlyValArgIleSerGlyAspIle 100  
 Db 241 CTGCGCACCCCTGCTGTTACAGCGCGCGGAAAGCGCTGCATCTCCGCTGACATC 300  
 Qy 101 AspLeuPheAlaArgGluAlaLysAlaProIleValAlaIleThrGlySerAsnAlaLys 120  
 Db 301 GATCTCTTCGGCCGCGAGCGGAAAGCGCGATCGTGGCCATCACCCTGTTCCAAACGCGAG 360  
 Qy 121 SerThrValThrThrLeuValGlyGluMetAlaValAlaAlaAspLysArgValAlaVal 140  
 Db 361 AGCACCGTGACCACTTGGTGGCGGAAATGGCGGTGGCGGACAAAGCGTGTCCGCGTC 420  
 Qy 141 GlyGlyAsnLeuGlyThrProAlaLeuAspLeuLeuAlaAspAspIleGluLeuTyrVal 160  
 Db 421 GCGCGCAACCTCGGACCCCGCGCTCGACCTGCTGGCCGACGATCGAGTGTACGTTG 480  
 Qy 161 LeuGluLeuSerSerPheGlnLeuGluThrCysAspArgLeuAsnAlaGluValAlaThr 180  
 Db 481 TTGGAGCTGTGAGCTTCCAGCTGGAAACCTGCGATCGCTCAACCGGAGGTGGCGACC 540  
 Qy 181 ValLeuAsnValSerGluAspHisMetAspArgTyrAspGlyMetAlaAspTyrHisLeu 200  
 Db 541 GTGCTGAACGTCAGCGAAGACCATATGATCGTACGACGCGCATGGCTGACTTACCACCTG 600  
 Qy 201 AlaLysHisArgIlePheArgGlyAlaArgGlnValValAlaAsnArgAlaAspAlaLeu 220  
 Db 601 GCCAAGACCGGATCTTCGCGGTGCGCGGAGTCTGGTGAATTCGCGCGCATGCCCTG 660  
 Qy 221 ThrArgProLeuLeuAlaAspThrValProCysTyrTrpSerPheGlyLeuAsnLysProAsp 240  
 Db 661 ACCGACCGCTGATGCGCGATACCGTGGCTGCTGCTGCTGGCGCTGAACAAAGCGGAC 720  
 Qy 241 PheLysAlaPheGlyLeuLeuIleGluAspGlyGlnLysTrpLeuAlaPheGlnPheAsp 260  
 Db 721 TTCAAGGCTTTCGGCTGATCGAGGAGACGCGCAGAGTGGCTGGCTTCCAGTTCGAC 780  
 Qy 261 LysLeuLeuProValGlyGluLeuLysIleArgGlyAlaHisAsnTyrSerAsnAlaLeu 280  
 Db 781 AAGTGTCTGGCGTGGCGAATCGATCGTGGCGCGCCACCACTATTTCCAAACGCGTC 840  
 Qy 281 AlaAlaLeuAlaLeuGlyHisAlaValGlyLeuProPheAspAlaMetLeuGlyAlaLeu 300  
 Db 841 GCCGCGTGGCGTGGCGCATCGGCTGCGCTGCGCTTCGACGCCATGCTCGGCGCGCTG 900  
 Qy 301 LysAlaPheSerGlyLeuAlaHisArgCysGlnTrpValArgGluArgGlnGlyValSer 320  
 Db 901 AAGCGTTCGCGCTGCTCATCGTCCAGTGGGTACGCGAGCGCGAGCGGCCGAT 960  
 Qy 321 TyrTyrAspAspSerLysAlaThrAsnValGlyAlaAlaLeuAlaAlaIleGluGlyLeu 340  
 Db 961 TACTACGACGATTCGAAGGCCACCAACGTCGCGCGCGCCCTGGCGGCGCATCGAGGGGCTG 1020  
 Qy 341 GlyAlaAspIleAspGlyLysLeuValLeuAlaGlyLysAspGlyLysGlyAlaAsp 360  
 Db 1021 GGTGCGGACATCGAGCGCAACGCTGCTGCTGCGCGCGGAGAGCGGCAAGCGGCCGAT 1080  
 Qy 361 PheHisAspLeuArgGluProValAlaAlaArgPheCysArgAlaValAlaLeuGlyArg 380  
 Db 1081 TTCCATGACCTCGCGAGCGCGCTCGCGCGCTTCGCGCGCGGTGGTACTGCTGGCGGT 1140  
 Qy 381 AspAlaGlyLeuIleAlaGlnAlaLeuGlyAsnAlaValProLeuValArgValAlaThr 400  
 Db 1141 GACCGCGGCTGATTTGCCAGCGACTGGGCAACGCGTACCGCTGGTGGCGCGCGCAACG 1200  
 Qy 401 LeuAspGluAlaValArgGlnAlaAlaGluLeuAlaArgGluGlyAspAlaValLeuLeu 420  
 Db 1201 CTGACCGAAGCGTCCGCGAGCGCGCGCTGCGCGCGGAGCGGATGCGGCTGCTGTG 1260  
 Qy 421 SerProAlaCysAlaSerLeuAspMetPheLysAsnPheGluGluArgGlyArgLeuPhe 440  
 Db 1261 TCGCGCGCTGCGCGAGCGCTGGACATGTTCAAGAACTTCGAAGAACGCGGACGCGCTGTC 1320  
 Qy 441 AlaLysAlaValGluGluLeuAla 448

```

Db      1321  GCCAAAGCCGTAGAGGAGCTAGCG 1344
RESULT 2
AAZ47131
ID      AAZ47131 standard; DNA; 1450 BP.
XX
AC      AAZ47131;
XX
DT      28-MAR-2000 (first entry)
XX
XX      Pseudomonas aeruginosa murD gene.
XX
DE      MurD; bacterial cell wall; biosynthesis; inhibitor; D-glutamate;
KW      UDP-N-acetylmuramyl-L-alanine precursor; ds.
XX
OS      Pseudomonas aeruginosa.
XX
FN      WO9961050-A1.
XX
PD      02-DEC-1999.
XX
PF      26-MAY-1999; 99WO-US11585.
XX
PR      29-MAY-1998; 98US-0087308.
XX
PA      (MERI ) MERCK & CO INC.
XX
PI      El-Sherbeini M, Azzolina B;
XX
DR      WPI; 2000-072548/06.
DR      P-PSDB; AAY56047.
XX
PT      New nucleic acid encoding the MurD protein of Pseudomonas aeruginosa,
PT      used to identify specific inhibitors -
XX
FS      Claim 4; Page 8-9; 35pp; English.
XX
CC      This sequence represents the MurD gene from Pseudomonas aeruginosa.
CC      MurD is an enzyme involved in biosynthesis of bacterial cell walls,
CC      so inhibiting it inhibits bacterial growth (by preventing
CC      MurD-catalyzed addition of D-glutamate to an alanyl residue in the
CC      UDP-N-acetylmuramyl-L-alanine precursor). Cells that express murD
CC      protein can be used: (a) to identify inhibitors of MurD, and (b) for
CC      recombinant production of MurD, which is used to raise specific
CC      antibodies (potential therapeutic inhibitors). Fragments of the murD
CC      gene can be used as probes and primers to detect MurD-encoding sequences
CC      (to identify polymorphic murD genes or to detect P. aeruginosa), also
CC      as antisense modulators. The murD inhibitors are useful for treating or
CC      preventing bacterial infections, including those caused by P. aeruginosa.
XX
SQ      Sequence 1450 BP; 227 A; 487 C; 487 G; 249 T; 0 other;

Alignment Scores:
Pred. No.:      0      Length:      1450
Score:          448.00  Matches:      448
Percent Similarity: 100.00%  Conservative: 0
Best Local Similarity: 100.00%  Mismatches: 0
Query Match:      100.00%  Indels: 0
DB:              21      Gaps: 0

US-09-701-229-2 (1-448) x AAZ47131 (1-1450)
Oy      1 MetSerLeuIleAlaSerAspHisPheArgIleValValGlyLeuGlyLysSerGlyMet 20
Db      51 ATGAGCTGATCGCTCCGACCACTTCGGATCGTTGTCGGCTCGGCAAGCGCGCATG 110
Oy      21 SerLeuValArgTyrLeuAlaArgArgGlyLeuProPheAlaValValAspThrArgGlu 40
Db      111 TCCCTGTGGCTACCTGGCGCGCGCTGCTTTCGCTTTCGCTGCTGATACCGAGAG 170
Oy      41 AsnProGluLeuAlaThrLeuArgAlaGlnTyrProGlnValGluValArgCysGly 60

```

```

171 AACCCGCGGAGCTGGCCACCCCTGCGTGCCAGTATCCGAGGTGGAAGTGCCTGGCGG 230
61 GluLeuAspAlaGluPheLeuCysSerAlaArgGluLeuTyrValSerProGlyLeuSer 80
231 GAACCTGACGCCGAGTTCTCTGCTCCGCCGCGGAACCTATGTGTCACGCCCGGCTTGTG 290
81 LeuArgThrProAlaLeuValGlnAlaAlaLysGlyValArgIleSerGlyAspIle 100
291 CTGGCACCCCTGCGCTGGTACAGCGCGCGGAAAGCGGTGCGCATCTCCGGTGCATC 350
101 AspLeuPheAlaArgGluAlaLysAlaProIleValAlaIleThrGlySerAsnAlaLys 120
351 GATCTCTTCGCCGCGAGCGGAAGGCCCGCATCGTCCCATCCCGTTCCAAGCGCAAG 410
121 SerThrValThrThrLeuValGlyGluMetAlaValAlaAspLysAlaValAlaVal 140
411 AGCACCTGTACCACTTGTGGGGAATGGGGTGGCCGCGACACCGTGTCCCGCTC 470
141 GlyGlyAsnLeuGlyThrProAlaLeuAspLeuAlaAspAspIleGluLeuTyrVal 160
471 GCGGCAACCTCGGCACCCCGCGCTCGACCTGCTGGCCGACGACATCGAGCTGTACGTG 530
161 LeuGluLeuSerSerPheGlnLeuGluThrCysAspArgLeuAsnAlaGluValAlaThr 180
531 TTGGAGCTGTGAGCTTCCAGCTTGAACCTGCGATCGCTCAACGCGGAGGTGGCGACC 590
181 ValLeuAsnValSerGluAspHisMetAspArgTyrAspGlyMetAlaAspTyrHisLeu 200
591 GTGCTGAACGTACGGAAGACCATATGAGTACGTCACGCGCATGGCTGACTACCACTG 650
201 AlaLysHisArgIlePheArgGlyAlaArgGlnValValValAsnArgAlaAspAlaLeu 220
651 GCCAAGCACCGGATCTTCCGCGTGCCCGCAGGTGCTGTGAATCGCGCGATGCGCTG 710
221 ThrArgProLeuIleAlaAspThrValProCysTrpSerPheGlyLeuAsnLysProAsp 240
711 ACCGCGCTGATCGCGCATACGTCGCTGCTGTGCTGCGGCTCGACAAAGCCGGAC 770
241 PheLysAlaPheGlyLeuIleGluGluAspGlyGlnLysTrpLeuAlaPheGlnPheAsp 260
771 TTCAAGCCTTTCGGCTGATCGAGGAAGACGCCAGAAAGTGGCTGGGCTTCCAGTTGAC 830
261 LysLeuLeuProValGlyGluLeuLysIleArgGlyAlaHisAsnTyrSerAsnAlaLeu 280
831 AAGTGTGTCGGTGGCGAACTGAGATCGCTGGCGCCCACTACTTCCAAGCGCTC 890
281 AlaAlaLeuAlaLeuGlyHisAlaValGlyLeuProPheAspAlaMetLeuGlyAlaLeu 300
891 GCGCGCTGGCGTGGGCGCATGCGCTGCGCTGCGCTGCGACGCCATGCTCGGCGCGCTG 950
301 LysAlaPheSerGlyLeuAlaHisArgCysGlnTrpValArgGluArgGlnGlyValSer 320
951 AAGCGCTTTCGGCGCTGCTATCGCTGCGTGGTACGCGAGCGCGAGCGGCGTGCAGC 1010
321 TyrTrpAspSerLysAlaThrAsnValGlyAlaAlaLeuAlaIleGluGlyLeu 340
1011 TACTACGACGATTCGAAGGCCACCAACGTCGCGCGCCCTTGGCGGCGATCGAGGGGCTG 1070
341 GlyAlaAspIleAspGlyLysLeuValLeuAlaGlyLysAspGlyLysValAsp 360
1071 GGTGCCGACATCGACGCAAGCTGTGCTGCTCCCGCGGAGAGCGCAAGCGGCGCGAT 1130
361 PheHisAspLeuArgGluProValAlaArgPheCysArgAlaValAlaLeuGlyArg 380
1131 TTCCATGACCTGCGGAGCGGTGCGCGCTTTCGCGGCGGGTGTGACTGCTTGGCGCT 1190
381 AspAlaGlyLeuIleAlaGlnAlaLeuGlyAsnAlaValProLeuValArgValAlaThr 400
1191 GACGCGGGTGTGATTGCCAGGCATGCGCAACCGGTACCGCTGGTGGCGCTGCGCAAG 1250
401 LeuAspGluAlaValArgGlnAlaAlaGluLeuAlaArgGluGlyAspAlaValLeuLeu 420
1251 CTGGACGAAGCAGTCCGCGAGCGCGAGTGGCCCGCGAAGCGATGCGGTGCTGTGTTG 1310

```

```
QY 421 SerProAlaCyAlaSerLeuAspMetPheLysAsnPheGluArgGlyArgLeuPhe 440
|||||
Db 1311 TCGCGGCGCTGCGGCGGACATGTTCAAGAACTTCAAGAAACGGCGGCGCTGTTTC 1370
|||||

QY 441 AlaLysAlaValGluGluLeuAla 448
|||||
Db 1371 GCCAAGCGGTAGAGGAGTAGCG 1394
|||||

RESULT 3
AAF28548/c
ID AAF28548 standard; DNA; 96109 BP.
XX
AC AAF28548;
XX
DT 04-APR-2001 (first entry)
XX
DE Genomic fragment #35.
XX
KW Genomic library; bacteria; human upper airway; otitis media; sinusitis;
KW bronchopulmonary; endocarditis; meningitis; ss.
XX
OS Moraxella catarrhalis.
XX
PN WO200078968-A2.
XX
PD 28-DEC-2000.
XX
PF 16-JUN-2000; 2000WO-US16649.
XX
PR 18-JUN-1999; 99US-0140121.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Lagace RE, Patterson C, Berg KL;
XX
PI WPI; 2001-041427/05.
XX
PT Genomic library for identifying diagnostic and therapeutic
PT compositions, and for identifying virulence factors, regulatory
PT elements and drug targets, comprises Moraxella catarrhalis nucleic
PT acids -
XX
PS Claim 1; Page 345-368; 545pp; English.
XX
CC The present invention relates to a Moraxella catarrhalis genomic library
CC comprising of a combination of 41 nucleic acid molecules (see
CC AAF28514-AAF28554). The library has a number of uses described in the
CC specification e.g. is useful for identifying diagnostic and therapeutic
CC compositions. M. catarrhalis (Branhamella catarrhalis) is a large
CC aerobic, gram-negative diplococcus, normally found among the bacterial
CC flora of human upper airways. M. catarrhalis is known to cause acute,
CC localised infections such as otitis media, sinusitis and bronchopulmonary
CC infection and life-threatening, systemic diseases including endocarditis
CC and meningitis.
XX
SQ Sequence 96109 BP; 28783 A; 18910 C; 20341 G; 28075 T; 0 other;

Alignment Scores:
Pred. No.: 1.02e-06 Length: 96109
Score: 19.00 Matches: 19
Percent Similarity: 100.00% Conservatve: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.24% Indels: 0
DB: 22 Gaps: 0

US-09-701-229-2 (1-448) x AAF28548 (1-96109)

QY 113 AlalleThrGlySerAsnAlaLysSerThrValThrThrLeuValGlyGluMetAla 131
|||||
Db 12400 GCCATACAGGCTCAATGCCAAAGACACTGTTACGACATTAGTAGGCGAGATGCCA 12344
|||||

RESULT 4
```

```
AAS89258
ID AAS89258 standard; cDNA; 1080 BP.
XX
AC AAS89258;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #25062.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
PI WPI; 2001-639362/73.
XX
DR P-PSDB; ABG25071.
XX
CC New isolated polynucleotide and encoded polypeptides, useful in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits and to assess
CC biodiversity.
XX
PS Claim 1; SEQ ID No 25062; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1080 BP; 267 A; 253 C; 299 G; 261 T; 0 other;

Alignment Scores:
Pred. No.: 0.00104 Length: 1080
Score: 14.00 Matches: 14
Percent Similarity: 100.00% Conservatve: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.12% Indels: 0
DB: 23 Gaps: 0

US-09-701-229-2 (1-448) x AAS89258 (1-1080)

QY 157 GluLeuTyrValLeuGluLeuSerSerPheGlnLeuGluThr 170
|||||
```



Db 22 GAACGTACGTGCTGGAACTGTCGAGCTTCAGCTCGAAACC 63

RESULT 5  
AAS89801  
ID AAS89801 standard; cDNA; 1080 BP.  
XX  
AC AAS89801;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE DNA encoding novel human diagnostic protein #25605.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI; 2001-639362/73.  
DR P-PSDB; ABG25614.  
XX  
New isolated polynucleotide and encoded polypeptides, useful in  
diagnostics, forensics, gene mapping, identification of mutations  
responsible for genetic disorders or other traits and to assess  
biodiversity  
XX  
PS Claim 1; SEQ ID No 25605; 103pp; English.  
XX  
The invention relates to isolated polynucleotide (I) and  
polypeptide (II) sequences. (I) is useful as hybridisation probes,  
polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
and gene mapping, and in recombinant production of (II). The  
polynucleotides are also used in diagnostics as expressed sequence tags  
for identifying expressed genes. (I) is useful in gene therapy techniques  
to restore normal activity of (II) or to treat disease states involving  
(II). (II) is useful for generating antibodies against it, detecting or  
quantitating a polypeptide in tissue, as molecular weight markers and as  
a food supplement. (II) and its binding partners are useful in medical  
imaging of sites expressing (II). (I) and (II) are useful for treating  
disorders involving aberrant protein expression or biological activity.  
The polypeptide and polynucleotide sequences have applications in  
diagnostics, forensics, gene mapping, identification of mutations  
responsible for genetic disorders or other traits to assess biodiversity  
and to produce other types of data and products dependent on DNA and  
amino acid sequences. AAS64197-AAS94564 represent novel human  
diagnostic coding sequences of the invention.  
Note: The sequence data for this patent did not appear in the printed  
specification, but was obtained in electronic format directly from WIPO  
at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 1080 BP; 267 A; 253 C; 299 G; 261 T; 0 other;  
Alignment Scores:  
Pred. No.: 0.00104 Length: 1080  
Score: 14.00 Matches: 14  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 3.12% Indels: 0  
Gaps: 0

US-09-701-229-2 (1-448) x AAS89801 (1-1080)

Oy 157 GluLeuTyrValLeuGluLeuSerSerPheGlnLeuGluThr 170  
|||||  
Db 22 GAACGTACGTGCTGGAACTGTCGAGCTTCAGCTCGAAACC 63

RESULT 6  
AAS90166  
ID AAS90166 standard; cDNA; 1080 BP.  
XX  
AC AAS90166;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE DNA encoding novel human diagnostic protein #25970.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI; 2001-639362/73.  
DR P-PSDB; ABG25979.  
XX  
New isolated polynucleotide and encoded polypeptides, useful in  
diagnostics, forensics, gene mapping, identification of mutations  
responsible for genetic disorders or other traits and to assess  
biodiversity  
XX  
PS Claim 1; SEQ ID No 25970; 103pp; English.  
XX  
The invention relates to isolated polynucleotide (I) and  
polypeptide (II) sequences. (I) is useful as hybridisation probes,  
polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
and gene mapping, and in recombinant production of (II). The  
polynucleotides are also used in diagnostics as expressed sequence tags  
for identifying expressed genes. (I) is useful in gene therapy techniques  
to restore normal activity of (II) or to treat disease states involving  
(II). (II) is useful for generating antibodies against it, detecting or  
quantitating a polypeptide in tissue, as molecular weight markers and as  
a food supplement. (II) and its binding partners are useful in medical  
imaging of sites expressing (II). (I) and (II) are useful for treating  
disorders involving aberrant protein expression or biological activity.  
The polypeptide and polynucleotide sequences have applications in  
diagnostics, forensics, gene mapping, identification of mutations  
responsible for genetic disorders or other traits to assess biodiversity  
and to produce other types of data and products dependent on DNA and  
amino acid sequences. AAS64197-AAS94564 represent novel human  
diagnostic coding sequences of the invention.  
Note: The sequence data for this patent did not appear in the printed  
specification, but was obtained in electronic format directly from WIPO  
at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 1080 BP; 267 A; 253 C; 299 G; 261 T; 0 other;  
Alignment Scores:  
Pred. No.: 0.00104 Length: 1080  
Score: 14.00 Matches: 14  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 3.12% Indels: 0

DB: 23 Gaps: 0

US-09-701-229-2 (1-448) x AAS90166 (1-1080)

QY 157 GluLeuTyrValLeuGluLeuSerSerPheGlnLeuGluThr 170  
|||||  
Db 22 GAACGTGACGCTGGAACGTGCGAGCTCCAGCTGGAACC 63

RESULT 7  
AAS53453  
ID AAS53453 standard; DNA; 1314 BP.  
XX AAS53453;  
XX  
XX 13-FEB-2002 (first entry)  
XX Haemophilus influenzae DNA for cellular proliferation protein #235.  
XX Antisense; ds; prokaryotic cellular proliferation gene;  
KW antibiotic; antibacterial; drug design.  
XX  
XX Haemophilus influenzae.  
XX  
XX WO200170955-A2.  
XX  
XX 27-SEP-2001.  
XX  
XX 21-MAR-2001; 2001WO-US09180.  
XX  
XX 21-MAR-2000; 2000US-191078P.  
XX 23-MAY-2000; 2000US-206848P.  
XX 26-MAY-2000; 2000US-207727P.  
XX 23-OCT-2000; 2000US-242578P.  
XX 27-NOV-2000; 2000US-253625P.  
XX 22-DEC-2000; 2000US-257931P.  
XX 16-FEB-2001; 2001US-269308P.  
XX (ELIT-) ELITRA PHARM INC.  
XX  
XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;  
XX Yamamoto RT, Xu HH;  
XX  
XX WPI; 2001-611495/70.  
XX P-PSDB; AAU35594.  
XX  
XX New polynucleotides for the identification and development of  
XX antibiotics, comprise sequences of antisense nucleic acids -  
XX  
XX Claim 27; Seq ID No 7090; 51lpp; English.  
XX  
XX The invention relates to antisense inhibitors of genes essential to  
XX prokaryotic cellular proliferation, their use in identifying the  
XX genes, their use in the discovery of novel antibiotics, the essential  
XX genes themselves and the encoded proteins. The prokaryotes used are  
XX Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella  
XX pneumoniae, pseudomonas aeruginosa and Enterococcus faecalis. The  
XX invention is also useful for the identification of potential new targets  
XX for antibiotic development. The antisense nucleic acids can also be used  
XX to identify proteins used in proliferation, to express these proteins,  
XX and to obtain antibodies capable of binding to the expressed proteins.  
XX The proteins can be used to screen compounds in rational drug discovery  
XX programmes. The antisense nucleic acid sequence is also useful to screen  
XX for homologous nucleic acids which are required for cell proliferation in  
XX a wide variety of organisms. The present sequence encodes an  
XX essential prokaryotic cellular proliferation protein.  
XX Note: The sequence data for this patent did not form part  
XX of the printed specification, but was obtained in electronic  
XX format directly from WIPO at  
XX ftp.wipo.int/pub/published\_pct\_sequences.

Sequence 1314 BP; 403 A; 214 C; 294 G; 403 T; 0 other;

Alignment Scores:

Pred. No.: 0.00125 Length: 1314  
Score: 14.00 Matches: 14  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 3.12% Indels: 0  
DB: 23 Gaps: 0

US-09-701-229-2 (1-448) x AAS53453 (1-1314)

QY 157 GluLeuTyrValLeuGluLeuSerSerPheGlnLeuGluThr 170  
|||||  
Db 457 GAACTTTATGTACTAGAGCTTTCTAGTTTTCAGCTTGAGACA 498

RESULT 8  
AAH81460  
ID AAH81460 standard; DNA; 1317 BP.  
XX AAH81460;  
XX  
XX 21-SEP-2001 (first entry)  
XX Escherichia coli protein encoding nucleotide sequence SEQ ID NO:259.  
XX Escherichia coli; identification; proliferation; microorganism;  
KW antimicrobial; antibacterial; antibiotic; gene therapy; diagnosis;  
KW bacterial growth inhibition; ds.  
XX Escherichia coli.  
XX OS  
XX WO200148209-A2.  
XX  
XX 05-JUL-2001.  
XX  
XX 19-DEC-2000; 2000WO-US34419.  
XX  
XX 23-DEC-1999; 99US-0173005.  
XX (ELIT-) ELITRA PHARM INC.  
XX  
XX Forsyth RA, Ohlsen KL, Zyskind JW;  
XX  
XX WPI; 2001-457376/49.  
XX P-PSDB; AAG98404.  
XX  
XX Novel nucleic acids encoding proteins required for Escherichia coli  
XX proliferation, useful for screening for antimicrobial agents -  
XX  
XX Claim 9; Page 388-389; 596pp; English.  
XX  
XX The present invention describes a purified or isolated nucleic acid  
XX sequence (I) consisting essentially of one of the 93 nucleotide sequences  
XX given in AAH81202 to AAH81294, where expression of the nucleic acid in a  
XX microorganism is capable of inhibiting proliferation of a microorganism.  
XX (I) have antibacterial and antibiotic activities, and can be used in  
XX gene therapy. Expression of (I) in a microorganism inhibits proliferation  
XX of the microorganism, and the manufactured antibiotic is useful for  
XX reducing the activity or level of a gene product required for  
XX proliferation of a microorganism in a subject, specifically humans. The  
XX nucleic acids that inhibit bacterial growth or proliferation can be used  
XX as antisense therapeutics for killing bacteria. In addition to  
XX therapeutic applications, the nucleic acid sequences complementary to  
XX sequences required for proliferation can be used as diagnostic tools.  
XX For example, nucleic acid probes complementary to proliferation-required  
XX sequences that are specific for particular species of microorganisms can  
XX be used as probes to identify particular microorganism species in  
XX clinical specimens. AAH81295 to AAH81487 encode the Escherichia coli  
XX proteins given in AAG98239 to AAG98431, and AAH81488 to AAH81491  
XX represent oligonucleotides, which are used in the exemplification of the  
XX present invention.

Sequence 1317 BP; 294 A; 323 C; 400 G; 300 T; 0 other;

Alignment Scores:

Pred. No.: 0.00126 Length: 1317  
 Score: 14.00 Matches: 14  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 3.12% Indels: 0  
 DB: 22 Gaps: 0

US-09-701-229-2 (1-448) x AAH81460 (1-1317)

QY 157 GluLeuTyrValLeuGluLeuSerPheGlnLeuGluThr 170  
 |||||  
 Db 457 GAACTGTACGTGCTGGAACTGTCGAGCTCCAGCTGGAAACC 498

RESULT 9  
 AAS52295  
 ID AAS52295 standard; DNA; 1317 BP.  
 XX  
 AC AAS52295;  
 XX  
 DT 13-FEB-2002 (first entry)  
 XX  
 DE E. coli DNA for cellular proliferation protein #17.  
 XX  
 KW Antisense; ds; prokaryotic cellular proliferation gene;  
 KW antibiotic; antibacterial; drug design.  
 XX

OS Escherichia coli.  
 XX  
 PN WO200170955-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 21-MAR-2001; 2001WO-US09180.  
 XX  
 PR 21-MAR-2000; 2000US-191078P.  
 PR 23-MAY-2000; 2000US-206848P.  
 PR 26-MAY-2000; 2000US-207727P.  
 PR 23-OCT-2000; 2000US-242578P.  
 PR 27-NOV-2000; 2000US-253625P.  
 PR 22-DEC-2000; 2000US-257931P.  
 PR 16-FEB-2001; 2001US-269308P.  
 XX  
 PA (ELIT-) ELITRA PHARM INC.

XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;  
 PI Yamamoto RT, Xu HH;  
 XX WPI: 2001-611495/70.  
 DR P-PSDB; AAU34436.

XX New polynucleotides for the identification and development of  
 PT antibiotics, comprise sequences of antisense nucleic acids -

XX Claim 27; Seq ID No 5932; 51lpp; English.

XX The invention relates to antisense inhibitors of genes essential to  
 CC prokaryotic cellular proliferation, their use in identifying the  
 CC genes, their use in the discovery of novel antibiotics, the essential  
 CC genes themselves and the encoded proteins. The prokaryotes used are  
 CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella  
 CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The  
 CC invention is also useful for the identification of potential new targets  
 CC for antibiotic development. The antisense nucleic acids can also be used  
 CC to identify proteins used in proliferation, to express these proteins,  
 CC and to obtain antibodies capable of binding to the expressed proteins.  
 CC The proteins can be used to screen compounds in rational drug discovery  
 CC programmes. The antisense nucleic acid sequence is also useful to screen  
 CC for homologous nucleic acids which are required for cell proliferation in  
 CC a wide variety of organisms. The present sequence encodes an  
 CC essential prokaryotic cellular proliferation protein.

CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at

CC ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 1317 BP; 294 A; 323 C; 400 G; 300 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 0.00126 Length: 1317  
 Score: 14.00 Matches: 14  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 3.12% Indels: 0  
 DB: 23 Gaps: 0

US-09-701-229-2 (1-448) x AAS52295 (1-1317)

QY 157 GluLeuTyrValLeuGluLeuSerPheGlnLeuGluThr 170  
 |||||  
 Db 457 GAACTGTACGTGCTGGAACTGTCGAGCTCCAGCTGGAAACC 498

RESULT 10  
 AAS56341  
 ID AAS56341 standard; DNA; 1317 BP.  
 XX  
 AC AAS56341;  
 XX  
 DT 13-FEB-2002 (first entry)  
 XX  
 DE Salmonella typhi DNA for cellular proliferation protein #374.  
 XX  
 KW Antisense; ds; prokaryotic cellular proliferation gene;  
 KW antibiotic; antibacterial; drug design.  
 XX

OS Salmonella typhi.  
 XX  
 PN WO200170955-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 21-MAR-2001; 2001WO-US09180.  
 XX  
 PR 21-MAR-2000; 2000US-191078P.  
 PR 23-MAY-2000; 2000US-206848P.  
 PR 26-MAY-2000; 2000US-207727P.  
 PR 23-OCT-2000; 2000US-242578P.  
 PR 27-NOV-2000; 2000US-253625P.  
 PR 22-DEC-2000; 2000US-257931P.  
 PR 16-FEB-2001; 2001US-269308P.

XX (ELIT-) ELITRA PHARM INC.  
 XX  
 PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;  
 PI Yamamoto RT, Xu HH;  
 XX WPI: 2001-611495/70.  
 DR P-PSDB; AAU38482.

XX New polynucleotides for the identification and development of  
 PT antibiotics, comprise sequences of antisense nucleic acids -

XX Claim 27; Seq ID No 9978; 51lpp; English.

XX The invention relates to antisense inhibitors of genes essential to  
 CC prokaryotic cellular proliferation, their use in identifying the  
 CC genes, their use in the discovery of novel antibiotics, the essential  
 CC genes themselves and the encoded proteins. The prokaryotes used are  
 CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella  
 CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The  
 CC invention is also useful for the identification of potential new targets  
 CC for antibiotic development. The antisense nucleic acids can also be used  
 CC to identify proteins used in proliferation, to express these proteins,  
 CC and to obtain antibodies capable of binding to the expressed proteins.  
 CC The proteins can be used to screen compounds in rational drug discovery  
 CC programmes. The antisense nucleic acid sequence is also useful to screen  
 CC for homologous nucleic acids which are required for cell proliferation in

CC a wide variety of organisms. The present sequence encodes an  
CC essential prokaryotic cellular proliferation protein.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 1317 BP; 258 A; 341 C; 427 G; 291 T; 0 other;

Alignment Scores:  
Pred. No.: 0.00126 Length: 1317  
Score: 14.00 Matches: 14  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 3.12% Indels: 0  
DB: 23 Gaps: 0

US-09-701-229-2 (1-448) x AAS56341 (1-1317)

QY 157 GluLeuTyrValLeuGluLeuSerSerPheGlnLeuGluThr 170  
AAS73012  
Db 457 GAATTGTACGTCTGGAATATATCCAGTTTCCCAATTGGAAC 498

RESULT 11  
AAS73012  
ID AAS73012 standard; cDNA; 2058 BP.

XX AAS73012;  
XX  
XX  
XX 13-FEB-2002 (first entry)  
XX  
XX DNA encoding novel human diagnostic protein #8816.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX Homo sapiens.  
XX WO200175067-A2.  
XX  
XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.  
XX  
XX 31-MAR-2000; 2000US-0540217.  
XX 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI: 2001-639362/73.  
XX P-PSDB; ABG08825.

XX New isolated polynucleotide and encoded polypeptides, useful in  
XX diagnostics, forensics, gene mapping, identification of mutations  
XX responsible for genetic disorders or other traits and to assess  
XX biodiversity

XX Claim 1; SEQ ID No 8816; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and  
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,  
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
XX and gene mapping, and in recombinant production of (II). The  
XX polynucleotides are also used in diagnostics as expressed sequence tags  
XX for identifying expressed genes. (I) is useful in gene therapy techniques  
XX to restore normal activity of (II) or to treat disease states involving  
XX quantitating a polypeptide in tissue, as molecular weight markers and as  
XX a food supplement. (II) and its binding partners are useful in medical  
XX imaging of sites expressing (II). (I) and (II) are useful for treating  
XX disorders involving aberrant protein expression or biological activity.

CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 2058 BP; 496 A; 526 C; 574 G; 462 T; 0 other;

Alignment Scores:  
Pred. No.: 0.00192 Length: 2058  
Score: 14.00 Matches: 14  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 3.12% Indels: 0  
DB: 23 Gaps: 0

US-09-701-229-2 (1-448) x AAS73012 (1-2058)

QY 157 GluLeuTyrValLeuGluLeuSerSerPheGlnLeuGluThr 170  
Db 1783 GAACGTACGTCTGGAACCTGTCGAGCTTCCAGCTGGAAC 1824

RESULT 12  
AAS69095

ID AAS69095 standard; cDNA; 3117 BP.

XX AAS69095;

XX 13-FEB-2002 (first entry)

XX DNA encoding novel human diagnostic protein #4899.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX Homo sapiens.

XX WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

XX 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI: 2001-639362/73.

XX P-PSDB; ABG04908.

XX New isolated polynucleotide and encoded polypeptides, useful in  
XX diagnostics, forensics, gene mapping, identification of mutations  
XX responsible for genetic disorders or other traits and to assess  
XX biodiversity

XX Claim 1; SEQ ID No 4899; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and  
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,  
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
XX and gene mapping, and in recombinant production of (II). The  
XX polynucleotides are also used in diagnostics as expressed sequence tags  
XX for identifying expressed genes. (I) is useful in gene therapy techniques  
XX to restore normal activity of (II) or to treat disease states involving  
XX quantitating a polypeptide in tissue, as molecular weight markers and as

CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. AAS64197-AAS94564 represent novel human  
 CC diagnostic coding sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 3117 BP; 749 A; 820 C; 856 G; 692 T; 0 other;

Alignment Scores:  
 Pred. No.: 0.00284 Length: 3117  
 Score: 14.00 Matches: 14  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 3.12% Indels: 0  
 DB: 23 Gaps: 0

US-09-701-229-2 (1-448) x AAS69095 (1-3117)

OY 157 GluLeuTyrValLeuGluLeuSerPheGlnLeuGluThr 170  
 |||||  
 DB 355 GAACGTACGTGCTGGAACGTGCGAGCTCCACGTGGAACC 396

RESULT 13

AAS89240  
 ID AAS89240 standard; cDNA; 3117 BP.

XX AC AAS89240;

XX DT 13-FEB-2002 (first entry)

XX DE DNA encoding novel human diagnostic protein #25044.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US08631.

XX PR 31-MAR-2000; 2000US-0540217.

XX PR 23-AUG-2000; 2000US-0649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX DR WPI; 2001-639362/73.

XX DR P-PSDB; ABG25053.

XX PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity

XX PS Claim 1; SEQ ID No 25044; 103pp; English.

XX CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques

CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. AAS64197-AAS94564 represent novel human  
 CC diagnostic coding sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 3117 BP; 749 A; 820 C; 856 G; 692 T; 0 other;

Alignment Scores:  
 Pred. No.: 0.00284 Length: 3117  
 Score: 14.00 Matches: 14  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 3.12% Indels: 0  
 DB: 23 Gaps: 0

US-09-701-229-2 (1-448) x AAS89240 (1-3117)

OY 157 GluLeuTyrValLeuGluLeuSerPheGlnLeuGluThr 170  
 |||||  
 DB 355 GAACGTACGTGCTGGAACGTGCGAGCTCCACGTGGAACC 396

RESULT 14

AAS94542  
 ID AAS94542 standard; cDNA; 4831 BP.

XX AC AAS94542;

XX DT 13-FEB-2002 (first entry)

XX DE DNA encoding novel human diagnostic protein #30346.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US08631.

XX PR 31-MAR-2000; 2000US-0540217.

XX PR 23-AUG-2000; 2000US-0649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX DR WPI; 2001-639362/73.

XX DR P-PSDB; ABG30355.

XX PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity

XX PS Claim 1; SEQ ID No 30346; 103pp; English.

XX CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques

CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. AAS64197-AAS94564 represent novel human  
 CC diagnostic coding sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 4831 BP; 1180 A; 1249 C; 1353 G; 1049 T; 0 other;

#### Alignment Scores:

Pred. No.:	0.00429	Length:	4831
Score:	14.00	Matches:	14
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	3.12%	Indels:	0
DB:	23	Gaps:	0

US-09-701-229-2 (1-448) x AAS94564 (1-4831)

QY 157 GluLeuTyrrValLeuGluLeuSerPheGlnLeuGluThr 170

DB 4556 GAACGTGACGTGCTGGAACCTGCGAGCTTCAGCTGGAACCC 4597

#### RESULT 15

AAS89816

ID AAS89816 standard; cDNA; 4942 BP.

XX AC AAS89816;

XX DT 13-FEB-2002 (first entry)

XX DE DNA encoding novel human diagnostic protein #25620.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US08631.

XX PR 31-MAR-2000; 2000US-0540217.

XX PR 23-AUG-2000; 2000US-0649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX DR WPI; 2001-639362/73.

XX DR P-PSDB; ASG25629.

XX PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity

XX PS Claim 1; SEQ ID No 25620; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. AAS64197-AAS94564 represent novel human  
 CC diagnostic coding sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 4942 BP; 1109 A; 1296 C; 1387 G; 1150 T; 0 other;

#### Alignment Scores:

Pred. No.:	0.00439	Length:	4942
Score:	14.00	Matches:	14
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	3.12%	Indels:	0
DB:	23	Gaps:	0

US-09-701-229-2 (1-448) x AAS89816 (1-4942)

QY 157 GluLeuTyrrValLeuGluLeuSerPheGlnLeuGluThr 170

DB 2179 GAACGTGACGTGCTGGAACCTGCGAGCTTCAGCTGGAACCC 2220

Search completed: August 14, 2003, 09:24:59  
 Job time : 395 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus.p2n model

Run on: August 14, 2003, 07:13:51 ; Search time 349 seconds  
(without alignments)  
2872.772 Million cell updates/sec

Title: US-09-701-229-2

Perfect score: 2260

Sequence: 1 MSIIASHDFRIVVGLSGM.....MFKNFEEGRGLFAKAVEELA 448

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1504479 seqs, 1118970152 residues

Total number of hits satisfying chosen parameters: 2626304

Minimum DB seq length: 25

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame\_plus.p2n.model -DEV=rlp  
-O=/cgn2\_1/USPTO.spool/US09701229/runat\_09082003\_161159\_4434/app\_query.fasta\_1.647  
-DB=PublishedApplications\_NA -QFMT=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1  
-LOOPCL=0 -LOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100  
-THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500  
-MINLEN=25 -MAXLEN=200000000  
-USER=US09701229.cgn2\_1.333 @runat\_09082003\_161159\_4434 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSFLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPEXT=10 -XGAPEXT=0.5 -FGAPOPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications\_NA:

1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq:\*  
10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq:\*  
12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*  
13: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq:\*  
14: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq:\*  
15: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*  
16: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*  
17: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
-----					

1	2260	100.0	1347	9	US-09-815-242-7912	Sequence 7912, Ap
2	989.5	43.8	1317	9	US-09-815-242-9978	Sequence 9978, Ap
3	985.5	43.6	1317	9	US-09-741-669-259	Sequence 259, App
4	985.5	43.6	1317	9	US-09-815-242-5932	Sequence 5932, Ap
5	879	38.9	1314	9	US-09-815-242-7090	Sequence 7090, Ap
6	879	38.9	1830121	14	US-10-329-960-1	Sequence 1, Appli
c	716	31.7	640681	10	US-09-790-988-1	Sequence 1, Appli
7	618.5	27.4	1401	14	US-10-156-761-6097	Sequence 1, Appli
8	618.5	27.4	9025608	14	US-10-156-761-6097	Sequence 1, Appli
9	592.5	26.2	1353	9	US-09-815-242-9535	Sequence 9535, Ap
10	586.5	26.0	1353	9	US-09-815-242-9300	Sequence 9300, Ap
11	534.5	23.7	1352	10	US-09-974-300-1685	Sequence 1685, Ap
12	525	23.2	1461	10	US-09-712-363-79	Sequence 79, Appl
13	519	23.0	1368	9	US-09-815-242-3936	Sequence 3936, Ap
14	519	23.0	1371	9	US-09-815-242-6539	Sequence 6539, Ap
15	493	21.8	1003	9	US-09-765-272-109	Sequence 109, App
16	493	21.8	1003	9	US-09-815-242-4419	Sequence 4419, Ap
17	470	20.8	1347	9	US-09-815-242-8250	Sequence 8250, Ap
18	470	20.8	1350	9	US-09-815-242-8695	Sequence 8695, Ap
19	470	20.8	1350	9	US-09-815-242-8695	Sequence 8695, Ap
c	463.5	20.5	3309400	10	US-09-738-626-1	Sequence 1, Appli
20	370.5	16.4	32768	10	US-09-070-927A-128	Sequence 128, App
21	329	14.6	1269	9	US-09-815-242-7232	Sequence 7232, Ap
22	324.5	14.4	1320	8	US-08-781-986A-80	Sequence 80, Appl
23	259.5	11.5	1392	14	US-10-156-761-2193	Sequence 1, Appli
24	259.5	11.5	9025608	14	US-10-156-761-2193	Sequence 1, Appli
c	25	10.3	1377	9	US-09-815-242-7914	Sequence 7914, Ap
26	232.5	9.8	1359	9	US-09-741-669-257	Sequence 257, App
27	220.5	9.8	1359	9	US-09-815-242-5930	Sequence 5930, Ap
28	220.5	9.8	1359	9	US-09-712-363-82	Sequence 82, Appl
29	213	9.4	1608	10	US-09-738-626-2370	Sequence 2370, Ap
30	203	9.0	468	10	US-09-738-626-2365	Sequence 2365, Ap
31	199.5	8.8	1359	9	US-09-815-242-9976	Sequence 9976, Ap
32	197	8.7	1428	9	US-09-815-242-7093	Sequence 7093, Ap
33	195.5	8.7	1443	9	US-09-815-242-7910	Sequence 7910, Ap
34	192	8.5	1455	9	US-09-815-242-9975	Sequence 9975, Ap
35	192	8.5	1551	10	US-09-738-626-2370	Sequence 2370, Ap
36	183	8.1	1458	10	US-09-738-626-2362	Sequence 2362, Ap
37	183	8.1	1485	10	US-09-712-363-76	Sequence 76, Appl
38	182.5	8.1	717	9	US-09-939-980-78	Sequence 2986, Ap
c	39	7.9	383	9	US-09-815-242-2986	Sequence 2986, Ap
40	177.5	7.9	1413	14	US-10-156-761-6095	Sequence 6095, Ap
41	175	7.7	1476	9	US-09-741-669-262	Sequence 262, App
42	175	7.7	1476	9	US-09-815-242-5935	Sequence 5935, Ap
43	174.5	7.7	1476	9	US-09-815-242-9981	Sequence 9981, Ap
44	174	7.7	1362	9	US-09-815-242-3822	Sequence 3822, Ap
45	174	7.7	1425	9	US-09-815-242-6676	Sequence 6676, Ap

#### ALIGNMENTS

#### RESULT 1

US-09-815-242-7912  
Sequence 7912, Application US/09815242  
Patent No. US20020061569A1  
GENERAL INFORMATION:  
APPLICANT: Hasebeck, Robert  
APPLICANT: Ohlsen, Karl L.  
APPLICANT: Zyskind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
FILE OF INVENTION: Prokaryotes  
FILE REFERENCE: ELITRA.011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
CURRENT FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26





### RESULT 3

```

US-09-741-669-259
; Sequence 259, Application US/09741669
; Patent No. US20020022718A1
; GENERAL INFORMATION:
; APPLICANT: Forsyth, R. Alllyn
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; TITLE OF INVENTION: Genes identified as r
; FILE REFERENCE: ELIIRA.009A
; CURRENT APPLICATION NUMBER: US/09/741,669
; CURRENT FILING DATE: 2000-12-19
; PRIORITY APPLICATION NUMBER: US 60/173005
; PRIOR FILING DATE: 1999-12-23
; NUMBER OF SEQ ID NOS: 481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 259
; LENGTH: 1317
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1317)
US-09-741-669-259

```

**Alignment Scores:**

Argument Scores:  
 Pred. NO.: 5.09e-107  
 Score: 985.50  
 Percent Similarity: 64.33%  
 Best Local Similarity: 51.24%  
 Query Match: 43.61%  
 DB: 9

US-09-701-229-2 (1-448) x US-09-741-669-259 (1-1317)

Qy	11	IleValValGlyLeuGlyLysSerGlyMetSerLeuValArgTyrLeuAlaArgArgGly	30
Db	28	GTCAATTATCGCGCTCACCGGGCTTTCTCGCTGGACTTTTCTCGCTCGCGGT	87
Qy	31	LeuProPheAlaValAspThrArgLysnProProGluLeuAlaThrLeuArgAla	50
Db	88	GTACAGCCGCGGTATGGATACGGGTATGACACCGCTGCCCTGGATGAATAA	141
Qy	51	GlnTyrProGln---ValGluValArgCysGlyGluLeuAspAlaGluPheLeuCysser	69
Db	142	-----CCGAGCCGTAGACGCCACACGGGCAGTCTGAATGATGAATGGCTGATGGCG	195
Qy	70	AlaArgGluLeuTyrValSerProGlyLeuSerLeuArgThrProAlaLeuValGlnAla	89
Db	196	GCAGATCTGATTGTCCAGTCCCGGTATTGCACCTGGCGCATCCATCCTTAAGCGCTGCC	255
Qy	90	AlaAlaLysGlyValArgIleSerGlyAspIleAspLeuPheAlaArgGluAlaLysAla	109
Db	256	GCTGATGCCGAATCGAATCGTGGCGATATCCAGCTGTTCTCTCCGGAAGCACAAACA	315
Qy	110	ProIleValAlaIleThrGlySerAsnAlaLysSerThrValThrThrLeuValGlyGlu	129
Db	316	CCGATTGTGCGGATTACCGGTTCTAACGCCAAAGACGCGTACCACCGCTAGTGGGTGAA	375
Qy	130	MetAlaValAlaAlaAspLysArgValAlaValGlyGlyAsnLeuGlyThrProAlaLeu	149
Db	376	ATGGCAAGCGCGGGGTTAACGTTGGTGTGGTGGCAATATTGCCCTGCCTGCGTTG	435
Qy	150	AspLeuLeuAlaAspAspIleGluTyrValLeuGluLeuLeuSerSerPheGlnLeuGlu	169
Db	436	ATGCTACTGGATGATGAGTGTGAAGTGTACGTGTGGAAGTGTCCGAGCTTCCAGCTGGAA	495
Qy	170	ThrCysAspArgLeuAsnAlaGluValAlaThrValLeuAsnValSerGluAspHisMet	189
Db	496	ACCACCTCCAGCTTACAGCGGTACGACGCGACCATCTTGAACGTGACTGAAGATCATATG	555
Qy	190	AspArgTyrAsp---GlyMetAlaAspTyrHisLeuAlaLysHisArgIlePheArgGly	208
Db	556	GATCGTATCCGTTTGGTTTACACAGPATCGTGCACAAAGCTGCATTTACGAAAAC	615
Qy	209	AlaArgGlnValValAlaAspAlaLeuThrArgProLeuIleAlaAspThr	228
Db	616	GGCAAGTTTGGCTGGTTAATGCTGATGCCCTTAACAATGCCGATCGCGGTGCGGAT	675
Qy	229	ValProCysTrpSerPheGlyLeuAsnLysProAspPheLysAlaPheGlyLeuIleGlu	248
Db	676	GAACGCTGCGTCAGCTTTGGCGTCAACATGGGTCACTATCAC-----CTGAATCAT	726
Qy	249	GluAspGlyGlnLysTrpLeuAlaPheGlnPheAspLysLeuLeuProValGlyGluLeu	268
Db	727	CACGAGGCGNAACCTGCTCGGGGTTAAAGGCCGAGAAAGTGTGAATGTGAAAGAGATG	786
Qy	269	LysIleArgGlyAlaHisAsnTyrSerAsnAlaLeuAlaAlaLeuGlyHisAla	288
Db	787	AAACTTCGGGCACCATAACTACACCAANTCGCTGGCGGCTGGCGCTGGCAGATGCT	846
Qy	289	ValGlyLeuProPheAspAlaMetLeuGlyAlaLeuLysAlaPheSerGlyLeuAlaHis	308
Db	847	GCAGGGTTACCGCGGCCAGCAGCCTGAAAGCGTTTACCACATTCACCTGCTGCCGCAT	906
Qy	309	ArgCysGlnTrpValArgGluArgGlnGlyValSerTyrAspAspSerLysAlaThr	328
Db	907	CCCTTTGAAGTTGTCTGGACATCAACGGGCTAGCTGGATTAAACGATTACGATTCGAAGCACC	966

Qy	329	AsnValGlnAlaLeuAlaAlaIleGluGlyLeuGlyAlaAspIleAspGlyLysLeu	348
Db	967	AACCTGCGCATACGGAAGCGCGCTGAATGGCCTG-----CACGTAGACGGCACACTG	1020
Qy	349	ValLeuLeuAlaGlyGlyAspGlyLysGlyAlaAspPheHisAspLeuArgGluProVal	368
Db	1021	CATTGTGTGGGTGGCGATGGTAAATCGCGGACTTT-----AGCCCACTG	1068
Qy	369	AlaArgPheCysArgAla-----ValValLeuLeuGlyArgAspAlaGlyLeu	384
Db	1069	GCGGTTCACCTGAATGCGGATACGTACGCTCTGTATTGTTTCGGTCGTGACGGCGCCGAC	1128
Qy	385	IleAlaGlnAlaLeuGlyAsnAlaValProLeuValArgValAlaThrLeuAspGluAla	404
Db	1129	CTGCG---GCGCTACGCCCGGAAGTG---GCAGAACAAACCGAAACTATGGAACAGCGC	1182
Qy	405	ValArgGlnAlaAlaGluLeuAlaAraGluGlyAspAlaValLeuLeuSerProAlaCys	424
Db	1183	ATGCGCTTGTGGCTCGCGTGTTCACCGCGGGGATATGTTCTGCTCTCCACAGCGCTGT	1242
Qy	425	AlaSerLeuAspMetPheLysAsnPheGluGlyArgLeuPheAlaLysAlaVal	444
Db	1243	GCCAGCCTTGATCAGTTCAGAACTTTGAACACACGAGGCAATGAGTTGCGCGTCTCGCG	1302
Qy	445	GluGluLeu	447
Db	1303	AAGGAGTTA	1311

## RESULT 4

```

US-09-815-242-5932
: Sequence 5932, Application US/09815242
: Patent NO. US20020061569A1
: GENERAL INFORMATION:
: APPLICANT: Haselbeck, Robert
: APPLICANT: Ohlsen, Kari L.
: APPLICANT: Zyskind, Judith W.
: APPLICANT: Wall, Daniel
: APPLICANT: Trawick, John D.
: APPLICANT: Carr, Grant J.
: APPLICANT: Yamamoto, Robert T.
: APPLICANT: Xu, H. Howard
: TITLE OF INVENTION: Identification of Essential Genes in
: TITLE OF INVENTION: Prokaryotes
: FILE REFERENCE: ELITRA.011A
: CURRENT APPLICATION NUMBER: US/09/815,242
: CURRENT FILING DATE: 2001-03-21
: PRIOR APPLICATION NUMBER: 60/191,078
: PRIOR FILING DATE: 2000-03-21
: PRIOR APPLICATION NUMBER: 60/206,848
: PRIOR FILING DATE: 2000-05-23
: PRIOR APPLICATION NUMBER: 60/207,727
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: 60/242,578
: PRIOR FILING DATE: 2000-10-23
: PRIOR APPLICATION NUMBER: 60/253,625
: PRIOR FILING DATE: 2000-11-27
: PRIOR APPLICATION NUMBER: 60/257,931
: PRIOR FILING DATE: 2000-12-22
: PRIOR APPLICATION NUMBER: 60/269,308
: PRIOR FILING DATE: 2001-02-16
: NUMBER OF SEQ ID NOS: 14110
: SOFTWARE: FastSEQ for Windows Version 4.0
: SEQ ID NO 5932
: LENGTH: 1317
: TYPE: DNA
: ORGANISM: Escherichia coli
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (1)...(1317)
US-09-815-242-5932

```

**Alignment Scores:**





OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (36636)..(36636)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (40808)..(40810)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (44416)..(44416)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (44905)..(44905)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (44975)..(44975)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (45593)..(45593)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (45732)..(45732)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (47036)..(47036)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (51334)..(51334)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (51602)..(51602)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (51786)..(51786)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (51805)..(51805)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (55369)..(55369)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (65309)..(65309)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (65313)..(65313)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (80024)..(80024)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (100091)..(100091)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (102696)..(102696)  
OTHER INFORMATION: n equals a, t, g or c

FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (105121)..(105121)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (107248)..(107248)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (117136)..(117136)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (119750)..(119750)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (119924)..(119924)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (120038)..(120038)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (121344)..(121344)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (122167)..(122167)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (122336)..(122336)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (131340)..(131340)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (131360)..(131360)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (139910)..(139910)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (140398)..(140398)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (142750)..(142750)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (145058)..(145058)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (145171)..(145171)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (145942)..(145942)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (147197)..(147197)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:



Qy 11 IleValValGlyLeuGlyLysSerGlyMetSerLeuValArgTyrLeuAlaArgArgGly 30  
Db 241263 TTAATTTAGGAATGGGTTAAACAGGAATATCTGCATTAATTTTTTTTAAAAAAGGA 241204  
Qy 31 LeuProPheAlaValValAspThrArgGluAsnProProGluLeuAlaThrLeuArgAla 50  
Db 241203 ATAAACCTAAATATTAGTAATCTAAACATCTTCAATTTTATTAATAAT----- 241150  
Qy 51 GlnTyrProGln---ValGluValArgCysGlyGluLeuAspAlaGluPheLeuCysSer 69  
Db 241149 -----CCTCAAAATATTGAATATAGTTAGGAAGTTTAGACCATCAGTGATCTAGAA 241096  
Qy 70 AlaArgGluLeuTyrValSerProGlyLeuSerLeuArgThrProAlaLeuValGlnAla 89  
Db 241095 TCAGACTTAATTTGTTATAAGTCCGGTATTCTTCAATTTAAACCTATTTTAATAAAGCA 241036  
Qy 90 AlaAlaLysGlyValArgIleSerGlyAspIleSerGlyAspLeuPheAlaArgGluAlaLysAla 109  
Db 241035 CGTTTGTAGTATTAGATTATTAGTATATTGAATCTTCAATTTTCTAGAGAAGTGACCTGT 240976  
Qy 110 ProIleValAlaIleThrGlySerAsnAlaLysSerThrValThrLeuValGlyGlu 129  
Db 240975 CCTATTATTCATTAACAGGACTAATGGAAGAAAGTACTGTAGTACCATGATTGAAGAA 240916  
Qy 130 MetAlaValAlaAlaAspLysArgValAlaValGlyGlyAsnLeuGlyThrProAlaLeu 149  
Db 240915 ATTCGAAAAAATCAGGATATAAGCTTTCTAGTGGTAAATAGTGTCCCGTACTA 240856  
Qy 150 AspleuLeuAlaAspAspIleGluLeuTyrValLeuGluLeuSerSerPheGlnLeuGlu 169  
Db 240855 GAAATCTGTGATAAAGAACGAGATTATACATAATGAAGTGTCTAGTTTTCACACTAGAA 240796  
Qy 170 ThrCysAspArgLeuAsnAlaGluValAlaThrValLeuAsnValSerGluAspHisMet 189  
Db 240795 AATACATTAAATTTAAATCTAAGATAGCAGTATTCTTAATATAGCGAAGATCATATC 240736  
Qy 190 AspArgTyr---AspGlyMetAlaAspTyrHisLeuAlaLysHisArgIlePheArgGly 208  
Db 240735 AATCGATACCAATGGATGCCAACATATAAAACACTAAATTTGTCTGTATAATCAAA 240676  
Qy 209 AlaArgGlnValValAsnArgAlaAspAlaLeuThrArgProLeuIleAlaAspThr 228  
Db 240675 GCAGAAATTTGTATTAATTAATCTAATGAAGATAGAAAGAAAGTCTTATTCAATAGTAAG 240616  
Qy 229 ValProCysTrp---SerPheGlyLeuAsnLysProAspPheLysAlaPheGlyLeuIle 247  
Db 240615 AATAAAAAATGGATTAGTTTGGACAAATAGAGTGTATTCGT-----ATTGT 240565  
Qy 248 GluGluAspGlyGlnLysTrpLeuAlaPheGlnPheAspLysLeuLeuProValGlyGlu 267  
Db 240564 TCTAAAAGCAATGACCCCTATTTTATTTTAAAAAATAAAAAATATAAATTAATTAATCTAGTAA 240505  
Qy 268 LeuLysIleArgGlyAlaHisAsnTyrSerAsnAlaLeuAlaLeuAlaLeuGlyHis 287  
Db 240504 ATATTATATATGATATCAATTAATCAATATATATATATATAGTCTCTTTAGCTATTTCAGAT 240445  
Qy 288 AlaValGlyLeuProPheAspAlaMetLeuGlyAlaLeuLysAlaPheSerGlyLeuAla 307  
Db 240444 GCAATGCAATTCCTAGAAATGACCAATTAATGCTACTTAAAGTTTCTCGAATTTACCA 240385  
Qy 308 HisArgCysGlnTrpValArgGluArgGlnGlyValSerTyrAspAspSerLysAla 327  
Db 240384 CATCGATTCAATTAATAAAAAATAAAGAGGAGTAGTGTGGATAAATCAATCTTCAATATCC 240325  
Qy 328 ThrAsnValGlyAlaAlaLeuAlaAlaIleGluGlyLeuGlyAlaAspIleAspGlyLys 347  
Db 240324 ACAAATGTAATAGTACTCAAGTTGCTTTAAATTTCTATA-----AAACAACCTGGGAACA 240271  
Qy 348 LeuValLeuLeuAlaGlyGlyAspGlyLysGlyAlaAspPheHisAspLeuArgGluPro 367  
Db 240270 ATAAGATTATTTAGCGCGGATAGTAATCTCAATTTTATATATATAAATAATTT 240211  
Qy 368 ValAlaArgPheCysArgAlaValValLeuLeuGlyArgAspAlaGlyLeuIleAlaGln 387

Db 240210 TTTAGAACTTTAAAAATAAAAAATTTACTGTTTTCGAAGAGATGCATTAATTTCTCAAAA 240151  
Qy 388 AlaLeuGlyAsnAlaValProLeuValArgValAlaThrLeuAspGluAlaValArgGln 407  
Db 240150 ATATGTGAAAA-----AATCTATTATGTGTGAAAAATTTAAAAAAGCAGTAAATTTTA 240097  
Qy 408 AlaAlaGluLeuAlaArgGluGlyAspAlaValLeuLeuSerProAlaLysAlaCysAlaSerLeu 427  
Db 240096 ATTTCAAAACACAGTCAAAATCAGGTGATACCGTTCTTTGCTCTCTGGTGCAGCAGCTTA 240037  
Qy 428 AspMetPheLysAsnPheGluGluArgGlyArgLeuPheAlaLysAlaValGluGluLeu 447  
Db 240036 GGTCATATTTCTTAATCTTTGAAGAGAGAGTAAATCTTTTATAAAATTAATAAAGCAATA 239977  
RESULT 8  
US-10-156-761-6097  
; Sequence 6097, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 6097  
; LENGTH: 1401  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(1401)  
US-10-156-761-6097  
Alignment Scores:  
Pred. No.: 2,14e-63 Length: 1401  
Score: 618.50 Matches: 180  
Percent Similarity: 50.21% Conservative: 55  
Best Local Similarity: 38.46% Mismatches: 188  
Query Match: 27.37% Indels: 45  
Gaps: 14  
US-09-701-229-2 (1-448) x US-10-156-761-6097 (1-1401)  
Qy 12 ValValGlyLeuGlyLysSerGlyMetSerLeuValArgTyrLeuAlaArgArgGlyLeu 31  
Db 31 GTCGCGGCTCGGCGTCTCGGTATCCGCGCGCCGCGTGTCTGTCACAGCGCTCGCGCG 90  
Qy 32 ProPheAlaValVal-----AspThrArgGluAsnProProGluLeuAlaThrLeu 48  
Db 91 GTCGTCACGGTCTCAACAGCAGCGCGCAGCAGCGCTCCCGCGCCAGCGCGCGATCTG 150  
Qy 49 ArgAlaGlnTyrProGlnValGluValArgCysGlyGluLeuAspAlaGluPheLeuCys 68  
Db 151 GAGCGGCTC-----GGCATCAGCGTGGCGCTCGGT-----GACGGCGCGACCTGCC 198  
Qy 69 SerAlaArgGluLeuTyrVal---SerProGlyLeuSerLeuArgThrProAlaLeuVal 87  
Db 199 GAGGCGCAGGCTCATCGTCACACACCGCGCTGGCAGCGCGCAGCAAGCCGCTGTTCGCC 258  
Qy 88 GlnAlaAlaLysGlyValArgIleSerGlyAspIleAspLeuPheAlaArg----- 105



Db 259 GCGGCGCGGAGCGGCGTCCCGGTCTGGGGCGGACGTGCAACTCGCCTGGCGGCTGGCG 318  
QY 106 ---GluAlaLysAlaProLeValAlaIleThrGlySerAsnAlaLysSerThrValThr 124  
Db 319 GCGCGCGGATCCGCGCCCTGGTGCCTGCGGTCACGGGCAACCAAGGACCAACCGCGT 378  
QY 125 ThrLeuValGlyCyluMetAlaValAlaAlaAspLysArgValAlaValGlyGlyAsnLeu 144  
Db 379 CAGATGCTCGCTCGATCTGAGCGCGCGGCGCTCGCACGCGCCGCTGGCGCAACATC 438  
QY 145 GlyThrProAlaLeuAsp-----LeuLeuAlaAspAspIleGluLeuTyrValLeuGlu 162  
Db 439 GCGGTCTGCTCTCGACCGGTCTCTCGCGGAGACGTACGACGCTCTCGCGTGCAGAA 498  
QY 163 LeuSerSerPheGlnLeuGluThrCysAspArgLeuAsnAlaGluValAlaThrValLeu 182  
Db 499 CTCCTCAGGTACCACTGACGCTGCGGCGGCGCTCTCTCGCGGCGCACCTCCGCCACCGCTCG 558  
QY 183 AsnValSerGluAspHisMetAspArgTyrAspGlyMetAlaAspTyrHisLeuAlaLys 202  
Db 559 ACATCGCGCGGACCACTCGACTGGCAGCGCTCCATGGAGGCGGTACACCGCCGCAAG 618  
QY 203 HisArgIlePheArgGlyAlaArg---GlnValValAlaAspArgAlaAspAlaLeuThr 221  
Db 619 GCGGTATCTACGAGGCAATCGGTGCGCTCGCTACAGCTGCGCGCAAGGCCACC 678  
QY 222 ArgProLeuLe-----AlaAspThrValPro---CysTrpSerPheGly-----Leu 236  
Db 679 GAGGACCTGTGCGCGCGGACGTCGAGGAGGCTGCGGCGCTCGCTTCACGCTG 738  
QY 237 AsnLysProAspPheLysAlaPheGlyLeuIleGlu-----248  
Db 739 GGCACGCGCGGACGTCCTCACTCGCGCTGCTGGAGGCGATCTGTGTCAGCGCGCTTC 798  
QY 249 ---GluAspGlyGlnLysTrpLeuAlaPheGlnPheAspLysLeuLeuProValGlyGlu 267  
Db 799 GTCGAGGACCGGCAAG-----ACGCGCGGAGCTGCGCGAGGTGCGGAC 846  
QY 268 LeuLysIleArgGlyAlaHisAsnTyrSerAsnAlaLeuAlaLeuAlaLeuGlyHis 287  
Db 847 GTGCATCCCGCGCGCGGCAACATCGCCCAACGCGCTTGGCGGCGCGCGCTGGCGCG 906  
QY 288 AlaValGlyLeuProPheAspAlaMetLeuGlyAlaLeuLysAlaPheSerGlyLeuAla 307  
Db 907 GCCTTCGGGGTGCCCGCTCGCGGTACGGGACGAGCTGCGGCGCTTCGCGCGGACGCG 966  
QY 308 HisArgCysGlnTrpValArgGluArgGlnGlyValSerTyrTyrAspSerLysAla 327  
Db 967 CACCGCATCGCGCAGCTCGCCGACGTGGAGGGGTACCTACATCGACGACTCCAGGCG 1026  
QY 328 ThrAsnValGlyAlaAlaLeuAlaIleGluGlyLeuGlyAlaAspIleAspGlyLys 347  
Db 1027 ACCAACACGACGCGCGGAGCCTCTTTGGCGGCTACGGGTCC-----1071  
QY 348 LeuValLeuLeuAlaGlyAspGlyLysGlyAlaAspPheHisAspLeuArgGluPro 367  
Db 1072 ATCGTCTGATCGCGGGGGTTCGCCAAGGCGCTCTCTCGACGAGCTGTGTGCCAAG 1131  
QY 368 ValAlaArgPheCysArgAlaValValLeuLeuGlyArgAspAlaGlyLeuAlaGln 387  
Db 1132 TCGGCACAGCGCTCGGGGTGTGTGTCTATCGTGGGTGCGATCGTGGCTGATCGTGA 1191  
QY 388 AlaLeuGlyAsnAlaValProLeuValArgValAlaThrLeuAspGlu-----403  
Db 1192 GCGCTTGGCGGACACGCGCGGAGGTACCGGTGCGACCTCGACCGGACGACACTGG 1251  
QY 404 -----AlaValArgGlnAlaAlaGluLeuAlaArgGluGlyAspAlaValLeu 419  
Db 1252 GCGATCGCGCGCTGTCCAGAGCGCGGGCGCTCGCGCTCGCGCGGACGACTGTG 1311  
QY 420 LeuSerProAlaCysAlaSerLeuAspMetPheLysAsnPheGluGluArgGlyArgLeu 439  
Db 1312 CTGGCGCGCGCTGTGCTCGATGGACATGTTCGCCCACTACACAAAGCGGCTGACCG 1371

QY 440 PheAlaLysAlaValGluGluLeu 447  
Db 1372 TTCGCGGAGCGGTTCGCGGATC 1395

## RESULT 9

US-10-156-761-1  
; Sequence 1, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 1  
; LENGTH: 9025608  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: misc.feature  
; LOCATION: (4187715)  
; OTHER INFORMATION: a, t, c, g, other or unknown  
US-10-156-761-1

Alignment Scores:  
Pred. No.: 7,39e-58 Length: 9025608  
Score: 618.50 Matches: 180  
Percent Similarity: 50.21% Conservative: 55  
Best Local Similarity: 38.46% Mismatches: 188  
Query Match: 27.37% Indels: 45  
DB: 14 Gaps: 14

US-09-701-229-2 (1-448) x US-10-156-761-1 (1-9025608)

QY 12 ValValGlyLeuGlyLysSerGlyMetSerLeuValArgTyrLeuAlaArgArgGlyLeu 31  
Db 7360919 GTCGCGGCGCTCGCGGTCTCCGGTATCCGCGGCGCGCGCTGCTGACGGCGCTCGCGCG 7360978  
QY 32 ProPheAlaValVal-----AspThrArgGluAsnProProGluLeuAlaThrLeu 48  
Db 7360979 GTCGTCACGGTCTCAACGACGCGGACGAGGCTCCCGGGCGCGGCGCGGATCTG 7361038  
QY 49 ArgAlaGlnTyrProGlnValGluValArgCysGlyGluLeuAspAlaGluPheLeuCys 68  
Db 7361039 GAGGCGCTC-----GGCATCACCGTGGCGCTCGGT-----GACGGCGCGACCGCTGCC 7361086  
QY 69 SerAlaArgGluLeuTyrVal---SerProGlyLeuSerLeuArgThrProAlaLeuVal 87  
Db 7361087 GAGGCGACGAGCTCATCTGTCACCGCGCTGGAGCGGACGCGGCGCGCTGTTCGCC 7361146  
QY 88 GlnAlaAlaAlaLysGlyValArgIleSerGlyAspIleAspLeuPheAlaArg-----105  
Db 7361147 GCGGCGCGCGGCGCGGCTCCCGGTCTGGGCGAGCTCGAAGCTCGCTGGCGGCTGCG 7361206  
QY 106 ---GluAlaLysAlaProIleValAlaIleThrGlySerAsnAlaLysSerThrValThr 124  
Db 7361207 GCGCGCGGATCGCGCGCTGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTG 7361266  
QY 125 ThrLeuValGlyGluMetAlaValAlaAspLysArgValAlaValGlyGlyAsnLeu 144  
Db 7361267 CAGATGCTCGCTCGATCTCTGACGCGCGCGCTGCGACGCGCGCGCTGCGCGCAACATC 7361326





```
QY 115 ThrGlySerAsnAlaLysSerThrValThrLeuValGlyGluMetAlaValAla 134
Db 352 ACAGGATCGAAGCGTGAAGCAACACACAGCATATGATGGGAAGTTTGTGACTGCTG 411
QY 135 AsPlysArgValAlaValGlyGlyAsnLeuGlyThrProAlaLeuAspLeu-----Leu 152
Db 412 GGGCAACATGGCTTTTATCAGGAATATCGGTCATCCTGCCAGTCAGGTGCTCAAA 471
QY 153 AlaAspPleGluLeuTyrValLeuGluLeuSerPheGlnLeuLuthrCysAsp 172
Db 472 GCATCAGATAAGGACACGCTGTTATGGAATTTCTTCTTCCAACTCATGGGTGTTCAA 531
QY 173 ArgLeuAsnAlaGluValAlaThrValLeuAsnValSerGluAspHisMetAspArgTyr 192
Db 532 GAATTCATCCAGAGATTCGGTTATTAACCACTCATGCCAACTCATATCGACTACCAT 591
QY 193 AspGlyMetAlaAspTyrHisLeuAlaLysHisArgIle-----PheArgGlyAla 209
Db 592 GGGTCATTTTCTGAATATAGCAGCAAGTGAATATCCAGAACAGATGACAGCAGCT 651
QY 210 ArgGlnValValAsnArgAlaAspAlaLeuThrArgProLeuIleAlaAspThr--- 228
Db 652 GATTTCTTCTTGAATTTAATCAAGACTTGGCAAAAGACTTGACTTCCAAGACAGAA 711
QY 229 -----ValProCysTrpSerPheGlyLeuAsnLysProAspPheLysAlaPheGly 245
Db 712 GCACGTGTTGTACCACTTTTCAACA-----CTGAAAGGTTGAT-----GGA 753
QY 246 LeuIleGluGluAspGlyGlnLysTrpLeuAlaPheGlnPheAspLysLeuLeuProVal 265
Db 754 GCTTATCTGGAGATGGTCAA-----CTCTACTCCGTTGGTGAAGTAGCTATGCGAGCG 807
QY 266 GlyGluLeuLysIleArgGlyAlaHisAsnTyrSerAsnAlaLeuAlaAlaLeu 285
Db 808 AATGAATCGGTGTTCCAGGTAGCCACAATGTGAAATGCCCTTGCAGCTATTGCTGTA 867
QY 286 GlyHisAlaValGlyLeuProPheAspAlaMetLeuGlyAlaLeuLysAlaPheSerGly 305
Db 868 GCCAGCTCGTGATGTGGACAAATCAACCACTCAAGAAACCTTTTCAGCCCTTCGGTGT 927
QY 306 LeuAlaHisArgCysGlnTrpValArgGluArgGlnGlyValSerTyrTyrAspAspSer 325
Db 928 GTCAAAACACCGTCTCCAGTTTGTGGATGACATCAAGGGTGTAAATCTCTATACACAGCT 987
QY 326 LysAlaThrAsnValGlyAlaAlaLeuAlaIleGluGlyLeuGlyAlaAspIleAsp 345
Db 988 AAATCAACTAATATCTTGGCTACTCAAAAGCCCTTATCAGGATTTGAC-----AAC 1038
QY 346 GlyLysLeuValLeuAlaGlyGlyAspGlyLysGlyAlaAspPheHisAspLeuArg 365
Db 1039 AGCAGGTGCTCTTATGTCAGGTGTTTGGACCGTGGCAATGAGTTGACCAATG--- 1095
QY 366 GluProValAlaArgPheCysArgAlaValValLeuLeuGlyArgAspAlaGlyLeuIle 385
Db 1096 GTGCCAGACATTAATCTGAGTCAAGAGATGGTCTATCCTGGGTCAATCTGCACAGCTGTC 1155
QY 386 AlaGlnAlaLeuGlyAsnAla---ValProLeuValArgValAlaThrLeuAspGluAla 404
Db 1156 AAACGGGACGACAGACAGCTGGTCTGCTTATGTGGAGGCGACAGATATTGCAGATGCG 1215
QY 405 ValArgGlnAlaAlaGluLeuAlaArgGluGlyAspAlaValLeuLeuLeuSerProAlaCys 424
Db 1216 ACCCGCAGGCTATGACCTTGGAGTCAAGAGATGGTCTTCTTATGCTCTGCTCCCAAT 1275
QY 425 AlaSerLeuAspMetPheLysAsnPheGluGluArgGlyArgLeuPheAlaLysAlaVal 444
Db 1276 GCCAGCTGGATATGATGCTTAACCTTGAAGTACGTGGCGACCTCTTCTATCGACACAGTA 1335
QY 445 GluGluLeu 447
Db 1336 GCGGAGTTA 1344
RESULT 11
```

```
US-09-815-242-9300
; Sequence 9300, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9300
; LENGTH: 1353
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; LOCATION: (1)...(1353)
; US-09-815-242-9300

Alignment Scores:
Pred. No.: 1,298-59 Length: 1353
Score: 586.50 Matches: 162
Percent Similarity: 52.92% Conservative: 83
Best Local Similarity: 34.95% Mismatches: 177
Query Match: 25.95% Indels: 41
DB: 9 Gaps: 14

US-09-701-229-2 (1-448) x US-09-815-242-9300 (1-1353)

QY 7 AspHisPheArg-----IleValValGlyLeuGlyLysSerGlyMetSerLeu 22
Db 13 GATCAATTAAATAAAGAAAGTTCTTTTAGTTGGCCCAAGTCGTGGTAATCTGCA 72
QY 23 ValArgTyrLeuAlaArgGlyLeuProPheAlaValAlaAspThrArg----- 39
Db 73 GCTCGTTTGTGGACAAGCTAGGTGCCATTTGTGACAGTAAATGATGGGAAACCTTTGAG 132
QY 40 GluAsnProGluLeuAlaThrLeuArgAlaGlnTyrProGlnValGluValArgCys 59
Db 133 GACATCCAGCTGCCCAAGTTTGTGTA-----GAAGGATCAAGTCTATTACA 183
QY 60 GlyGlu-----LeuAspAlaGluPheLeuCysSerAlaArgGluLeuTyr 74
Db 184 GGTGCCCATCTTTGGAACCTCTTGATGAAGAGTTTGCCTTATGGTGAAA----- 234
QY 75 ValSerProGlyLeuSerLeuArgThrProAlaLeuValGlnAlaAlaLysGlyVal 94
Db 235 ---AATCCAGGTATCCCTTACAAACAATCCCATGATTGAAAGGCTTTGCCCAAGAGAAT 291
QY 95 ArgIleSerGlyAspIleAspLeuPheAlaArgGluAlaLysAlaProIleValAlaIle 114
```

```
Db 292 CCAGTCTTGAGTGGAGTGGCTTATTGTTTTCAGAACCCGATTATTGGTATC 351
Qy 115 ThrGlySerAsnAlaLysSerThrValThrThrLeuValGlyGluMetAlaValAla 134
Db 352 ACAGATCGAAGCGTGAACACACACACGACTATGATGGGGAAGTTTGGACTGCTGCT 411
Qy 135 AspLysArgValAlaValGlyGlyAsnLeuGlyThrProAlaLeuAspLeu-----Leu 152
Db 412 GGGCAACATGCTTTTATCATCGGAATATCGCTATCTCCGACGCTGCTCAATA 471
Qy 153 AlaAspAspIleGluLeuThrValLeuGluLeuSerPheGlnLeuGluThrCysAsp 172
Db 472 GCATCAGATAGGACGCTGTTGATGGAATCTTCTTTCCAACTCATCGGTGTTCAA 531
Qy 173 ArgLeuAsnAlaGluValAlaThrValLeuAsnValSerGluAspHisMetAspArgTyr 192
Db 532 GAATTCATCCAGAGATGCGGTTATACCAACCTCATGCCAATCATATCGATACAT 591
Qy 193 AspGlyMetAlaAspTyrHisLeuAlaLysHisArgIle-----PheArgGlyAla 209
Db 592 GGGTCATTTCTGGAATATGATGACCAAGTGAATATCCAGAACAGATGACAGCAGCT 651
Qy 210 ArgGlnValValAlaAsnArgAlaAspAlaLeuThrArgProLeuIleAlaAspThr--- 228
Db 652 GATTTCCTGCTGTTGAATCTTAATCAAGACTTGGCAAAAGACTTGACTTCCAAAGACAA 711
Qy 229 -----ValProCysTyrPheGlyLeuAsnLysProAspPheLysAlaPheGly 245
Db 712 GCCACTGTTGATCACTTTTCAACA-----CTTGAAGAAGTTGAT-----GGA 753
Qy 246 LeuIleGluGluAspGlyGlnLysTrpLeuAlaPheGlnPheAspLysLeuLeuProVal 265
Db 754 GCTTATCTGGAAGATGTTCAA-----CTCTACTTCGCTGGTGAAGTATCATGCGACGC 807
Qy 266 GlyGluLeuLysIleArgGlyAlaHisAsnTyrSerAsnAlaLeuAlaAlaLeuAlaLeu 285
Db 808 AATGAATCGGTGTTCCAGTAGGACCAATGTGGAATGCGCTGGGACATTTGCTGTA 867
Qy 286 GlyHisAlaValGlyLeuProPheAspAlaMetLeuGlyAlaLeuLysAlaPheSerGly 305
Db 868 GCCAGCTCTCGTATGGACCAATCAACCATCAAGGAACCTCTTCAGCTTCGCTGGT 927
Qy 306 LeuAlaHisArgCysGlnTrpValArgGluArgGlnGlyValSerTyrTyrAspSer 325
Db 928 GTCAAAACACCTCTCCAGTTTGTGATGCATCAAGGGTGTAAATTTCTATAACGACAGT 987
Qy 326 LysAlaThrAsnValGlyAlaAlaLeuAlaIleGluGlyLeuGlyAlaAspIleAsp 345
Db 988 AAATCAACTAATATCTTGCTTACTCAAAAGCCCTTATCAGGATTTGAC-----AAC 1038
Qy 346 GlyLysLeuValLeuAlaGlyGlyAspGlyLysGlyAlaAspPheHisAspLeuArg 365
Db 1039 AGCAAGTCTGCTGATTCAGGTGGTGGTGGACCGTGGCAATGATTTGACCAATG--- 1095
Qy 366 GluProValAlaAlaArgPheCysArgAlaValValLeuLeuGlyArgAspAlaGlyLeuIle 385
Db 1096 GTGCCAGACATTAAGTGGACCTCAAGAAGATGCTCATCTCGGTGCTATCTGCAGACGCTC 1155
Qy 386 AlaGlnAlaLeuGlyLysAla-----ValProLeuValArgValAlaThrLeuAspGluAla 404
Db 1156 AAAGGGGCGACAGCAAGGCTGGTGTCTTATGGAGGGCGACAGATATTCAGATGCG 1215
Qy 405 ValArgGlnAlaAlaGluLeuAlaArgGluGlyAspAlaValLeuLeuSerProAlaCys 424
Db 1216 ACCGCAAGGCTATGAGCTGCGACTCAAGGAGATGTGGTCTCTTAGTCTCTGCGCAAT 1275
Qy 425 AlaSerLeuAspMetPheLysAsnPheGluGluArgGlyArgLeuPheAlaLysAlaVal 444
Db 1276 GCCAGCTGGGATATGATGCTAACTTTGAAGTACGTGGCGACCTCTTTATCGACACAGTA 1335
Qy 445 GluGluLeu 447
|||||
```

```
Db 1336 GCGGAGTTA 1344
RESULT 12
US-09-974-300-1685
; Sequence 1685, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; TITLE OF INVENTION: Expression
; FILE REFERENCE: 10085-500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1685
; LENGTH: 1352
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-1685
Alignment Scores:
Pred. No.: 1,92e-53 Length: 1352
Score: 534.50 Matches: 155
Percent Similarity: 50.32% Conservative: 78
Best Local Similarity: 33.48% Mismatches: 179
Query Match: 23.65% Indels: 51
DB: 10 Gaps: 16
US-09-701-229-2 (1-448) x US-09-974-300-1685 (1-1352)
Qy 11 IleValValGlyLeuGlyLysSerGlyMetSer-----Leu 22
Db 37 CTCCTACTGGGGCTGGCGAAAGGGATGATGCTCGCTCGATCTTCATGAGAAATGC 96
Qy 23 ValArgTyrLeuAlaArgArgGlyLeuProPheAlaValAlaValAspThrArgGluAsnPro 42
Db 97 GTAAAGGTGTCGTGAATGACGAGTGCCTTTT-----GAGGAAATGAA 141
Qy 43 ProGluLeuAlaThrLeuArgAlaGlnTyrProGlnValGluValArgCysGlyGluLeu 62
Db 142 CCT-----GCGCGATTTCTCGCGAA---AAAGGGTGGAGTCTGCGGCTCTCAT 192
Qy 63 AspAlaGluPhe-----LeuCysSerAlaArgGluLeuTyrValSerProGlyLeuSer 80
Db 193 CCGACGGAACTGTTTGTATCTTCATTCGATCGACATTTGATTAATAATCCGGCATCCGC 252
Qy 81 LeuArgThrProAlaLeuValGlnAlaAlaLysGlyValArgIleSerGlyAspIle 100
Db 253 TATGAAAACGTGATGGTTCGAAGACGCTGAAGCGGGTATCCCGGTCTGCAGCGAGTT 312
Qy 101 AspLeuPheAlaArgGluAlaLysAlaProIleValAlaIleThrGlySerAsnAlaLys 120
Db 313 GAATCGCTCATCATCTGCAGATGCTCCGTTTCATCGGCATTCACGGTTCAACACGGAAG 372
Qy 121 SerThrValThrLeuValGlyGluMetAlaValAlaAlaAspLysArgValAlaVal 140
Db 373 ACGACACGCAACGCTTGTATTGAATGCTGAAGCGCGATTCGAAAAAAGCGCTCGTT 432
Qy 141 GlyGlyAsnLeuGlyThrProAlaLeuAspLeu-----AlaAspAspIleGluLeu 158
Db 433 GCTGGGAATATCGGTGCGCGGCAAGGAGGATCGCAATCATCGACGCGGACGAGCAATGG 492
Qy 159 TyrValLeuGluLeuSerSerPheGlnLeuGluThrCysAspArgLeuAsnAlaGluVal 178
Db 493 ATCTGTGACAGCTGTCGAGTTTTCAGCTCATGGGACTTATCAATTCAGACCGCAAGATT 552
Qy 179 AlaThrValLeuAsnValSerGluAspHisMetAspArgTyrAspGlyMetAlaAspTyr 198
```

```
Db 553 GGATTATTATAACGCTATTGATGCCCATCTCGATTATACCATTCGCGCAAAATAT 612
QY 199 HisLeuAlaLysHisArgIlePheArgGlyAlaArgGln-----ValValValAsn 215
Db 613 GACCTTGCANACAGCAGGTCTATCAATCAATCGGAACAGACGCTGCCGCTGTGCAAT 672
QY 216 ArgAlaAspAlaLeuThrArgProLeuIleAlaAspThrValProCysTrpSerPheGly 235
Db 673 TTGGACGAT-----GAGACAGTCGTCGCGCTCGCGAGTGT-----708
QY 236 LeuAsnLysProAspPheLysAlaPhe-----GlyLeuIle 247
Db 709 ---TCAAAAGCGGAGAGGTGATTTTTCGTCGCGCGGACACTTGAGCGCGGAGCGTGC 765
QY 248 GluGluAspGlyGlnLysTrpLeuAlaPheGlnPheAspLysLeuLeuProValGlyGlu 267
Db 766 GTCAAGACGCGCA-----ATCATGTTTAAACGCGCAACCGGTCAATCGCTGGAAGAC 819
QY 268 LeuLysIleArgGlyAlaHisAsnTrpSerAsnAlaLeuAlaAlaLeuAlaGlyHis 287
Db 820 GTCGTTTTCGCGGAGAGCACAATTTGGAGAACATTTTGGCGCCCTTTCATCGTTAA 879
QY 288 AlaValGlyLeuProPheAspAlaMetLeuGlyAlaLeuLysAlaPheSerGlyLeuAla 307
Db 880 ACAGCAGCGGTTCGACGCGCGCTCGTAAAGTCTGACGAGCTTTACAGGTGTTAAG 939
QY 308 HisArgCysGlnTrpValArgGluArgGlnGlyValSerTyrTyrAspAspSerLysAla 327
Db 940 CACAGGATCATACGTCGCGGACGATCAAAACAGACTGTTTACAAATCAGCAGCAAGCG 999
QY 328 ThrAsnValGlyAlaAlaLeuAlaIleGluGlyLeuGlyAlaAspIleAspGlyLys 347
Db 1000 ACAACATT-----CTTGGCAGAAAAAAGCGCTGCGGCC-----TTTCAAAAGCGC 1047
QY 348 LeuValLeuAlaGlyGlyAspGlyLysGlyAlaAspPheHisAspLeuArgGluPro 367
Db 1048 GTCATTTTCTGCGAGGGGGCTTGACCGCGGAATGAATTTGATGAACATAAG---CCG 1104
QY 368 ValAlaArgPheCysArgAlaValValLeuLeuGlyArgAspAla-----GlyLeu 384
Db 1105 CATATGCTTTTGTAAAGCGGTGATCACTTCGCGGAGACCGCGCGGAGTTGAGAAG 1164
QY 385 IleAlaGlnAlaLeuGlyAsnAlaValProLeuValArgValAlaThrLeuAspGluAla 404
Db 1165 CTGGCGGAGAAATGGGA---ATCAACAGGTAAACGTGTGATAATTTGTAACAAGCA 1221
QY 405 ValArgGlnAlaAlaGluLeuAlaArgGluGlyAspAlaValLeuLeuSerProAlaCys 424
Db 1222 GCAACTGCGCGCTTCAGCTCTCAGACGAGGAGATGTCATCTCTGTCGCGCGCTGC 1281
QY 425 AlaSerLeuAspMetPheLysAsnPheGluGlyArgGlyArgLeuPheAlaLysAlaVal 444
Db 1282 GCAAGCTGGGATCAGTACAAACATTTGAAGAACGTGGTGACATGTTTGTAAACGCGTG 1341
QY 445 GluGluLeu 447
Db 1342 CATATGCTT 1350
```

## RESULT 13

```
US-09-712-363-79
; Sequence 79, Application US/09712363
; Patent No. US20020164588A1
; GENERAL INFORMATION:
; APPLICANT: Eisenberg, David
; APPLICANT: Rotstein, Sergio H.
; APPLICANT: Marcotte, Edward M.
; TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND
; INTERACTIONS OF PROTEINS BY COMPARATIVE ANALYSIS
; FILE REFERENCE: 07419-032001
; CURRENT APPLICATION NUMBER: US/09/712,363
; CURRENT FILING DATE: 2000-11-13
; PRIOR APPLICATION NUMBER: PCT/US00/02246
```

```
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/179,531
; PRIOR FILING DATE: 2000-02-01
; PRIOR APPLICATION NUMBER: 60/117,844
; PRIOR FILING DATE: 1999-01-29
; PRIOR APPLICATION NUMBER: 60/118,206,
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: 60/126,593
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 60/134,093
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/134,092
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/165,124
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/165,086
; PRIOR FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 292
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 79
; LENGTH: 1461
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
US-09-712-363-79

Alignment Scores:
Pred. No.: 2,89e-52 Length: 1461
Score: 525.00 Matches: 167
Percent Similarity: 45.36% Conservative: 62
Best Local Similarity: 33.81% Mismatches: 182
Query Match: 23.23% Indels: 84
DB: 10 Gaps: 14

US-09-701-229-2 (1-448) x US-09-712-363-79 (1-1461)
```

```
QY 11 IleValValGlyLeuGlyLysSerGlyMetSerLeuValArgTyrLeuAlaArgGly 30
Db 34 TTGGTAGCGGTGGCGGGTGACCGGTCAAGCGGTGGCGGTGCTGCTACTCGGTTTGGT 93
QY 31 LeuProPheAlaValAlaAspThrArgGluAsnProProGluLeuAlaThrLeuArgAla 50
Db 94 GCGACGCCGACGCTGTCGAC-----GACGATCCGGTCTATGCTCGCA-----135
QY 51 GlnTyrProGlnValGluValArgCysGlyGluLeuAspAlaGluPheLeuCysSer-- 69
Db 136 -----CCGACGCCGAGACGTGGGTGCTGCGACCGTGTAGTTCCTCGACGCGGTGACGAC 189
QY 70 -----AlaArgGluLeuTyrValSerProGlyLeuSerLeuArgThrProAlaLe 86
Db 190 ATAAACGGGTATGCGGTG--GTGGTCCGCTCCCGCTTCTCGCCGCAACCCCGCTACT 248
QY 86 uValGlnAlaAlaLysGlyValArgIleSerGlyAspIleAspLeuPheAlaArgG1 106
Db 249 GCGCCGCGCGCGCGCGGGGTGCGGATCGCGGTGACGTGGAGTGTAGCTGCGGCGGT 308
QY 106 uAlaLysAlaProIle-----ValAlaIleThrGlySerAsnAl 119
Db 309 AGACGACGCGGCTGCTACGAGCCGCGCGAGCTGGTGTGGTGTGACCGGCAACCGG 368
QY 119 aLysSerThrValThrLeuValGlyGluMetAlaValAlaAlaAspLysArgValAl 139
Db 369 CAAGACCAACGACGATGATGTCACGCCATGTCGTCGCGGTGGCGCGCGCGCGCT 428
QY 139 aValGlyGlyAsnLeuGlyThrProAlaLeuAspLeuAlaAspAspIleGluLeuTy 159
Db 429 GCTGTGCGGCATATCGCAGTCGCGGTGCTGTGATGCTGGGACGACCGCGCGAGTGT 488
QY 159 rValLeuGluLeuSerSerPheGlnLeuGluThrCysAspArgLeuAsnAlaGluValAl 179
Db 489 GCGCGTGGAGTTGTCCAGTTTCCAGTTCGACATGCGGTGGCGCGGTGCGGTGGCGG 548
QY 179 aThrValLeuAsnValSerGluAspHisMetAspArgTyrAspGlyMetAlaAspTyrH1 199
```

```
Db 549 CGCGGTCTCAACATTCCGAGACACCTGGACTGGCATGCCAGTATCCGCAATACAC 608
Qy 199 sLeuAlaLysHisArgIlePheArgGlyAlaArgGlnValValAlaAspAla 219
Db 609 CGCGCCCAAGCCCGGTGTGACCGCGGG-----GTAGCGGTGGCCGGGTGATGA 662
Qy 219 aLeuThrArgProLeuIleAlaAspThrValProCysTrp-----SerPheGlyLe 236
Db 663 CAGCCGAGCGCGGCACACTGTGACGGCTTCCGCGCGAGTCCGCGGTCCGGCT 722
Qy 236 uAsnLysProAspPheLysAlaPheGly-----LeuIleGluAspG1 251
Db 723 CGCGAGCCCGCGCGGAACTGGCGGTGGCGAGCCACCTGGTCGATCGC----- 777
Qy 251 yGlnLysTrpLeuAlaPheGlnPheAsp---LysLeuLeuProValGlyGluLeuLysI1 270
Db 778 -----GCCTCTCCGACGACTTACGCTGCTCCGGTCCGCTCGATACCGGT 824
Qy 270 eArgGlyAlaHisAsnTyrSerAsnAlaLeuAlaLeuAlaLeuGlyHisAlaValG1 290
Db 825 GCCAGTTCGGTGGCTGGCTGACGCCCTGGCGCGCGCTGGCGCGCTCGGTGCG 884
Qy 290 yLeuProPheAspAlaMetLeuGlyAlaLeuLysAlaPheSerGlyLeuAlaHisArgCy 310
Db 885 GTGCGCCCGCGGTGCGATGCCGCGCGGTTCAGTTCGATGGCGCGCACCGCGC 944
Qy 310 sGlnTrpValArgGluArgGlnGlyValSerTyrTyrAspSerLysAlaThrAsnVa 330
Db 945 CGAGGTGGTGGCGGTTCGCCGCGCATCCTACCTAGTGGAGCATCCAGGCCACCAACC 1004
Qy 330 lGlyAlaAlaLeuAlaAlaIleGluGlyLeuGlyAlaAspIleAspGlyLysLeuValLe 350
Db 1005 GCAGCGCGCGGTGGCTGGCTGGCTGCTGCATAC-----CCGAGGTGGTATG 1049
Qy 350 uLeuAlaGlyGlyAspGlyLysGlyAlaAspPheHisAspLeuArgGluProValAlaAr 370
Db 1050 GATCGCGGTGGCTGCTCAAGCGCGGTGCTTCAC-----GCCAGGTGGCGCGAT 1103
Qy 370 gPheCysArgAlaVal-----ValLeuLeuGlyArgAspAlaGlyLeuIleAlaGlnAl 388
Db 1104 GCGTTCGCGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1163
Qy 388 aLeuGlyAsn-----AlaValProLeuValArgVal----- 398
Db 1164 GTTATCACGACACGCGCGCGATGTCCAGTTCGTTACAGTTCGTCAGGCGGAGGATAC 1223
Qy 398 ----- 398
Db 1224 TATGCTCGGACTGTGAGTTCCTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1283
Qy 399 -----AlaThrLeuAspGluAlaValArgGlnAlaAlaGluLe 411
Db 1284 AGCGGTGAGACCGTGGCGTGGCGTGGCGTGGCGTGGCGTGGCGTGGCGTGGCGTGGCG 1343
Qy 411 uAlaArgGluGlyAspAlaValLeuLeuSerProAlaCysAlaSerLeuAspMetPheLy 431
Db 1344 GGCCCAACCGGTGACACCGTGTGCTGGCAGCGCGCGCTCATTCGACCATTCAC 1403
Qy 431 sAsnPheGluGluArgGlyArgLeuPheAlaLysAlaVal 444
Db 1404 CGGTTATGCCGACCGCGCGGAGGATTCGCGCGCGGT 1443
```

## RESULT 14

```
US-09-815-242-3936
; Sequence 3936, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlson, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
```

```
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: XU, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011a
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR FILING DATE: 2000-03-21
; PRIOR FILING DATE: 2000-03-21
; PRIOR FILING DATE: 2000-05-23
; PRIOR FILING DATE: 2000-05-23
; PRIOR FILING DATE: 2000-05-26
; PRIOR FILING DATE: 2000-05-26
; PRIOR FILING DATE: 2000-10-23
; PRIOR FILING DATE: 2000-11-27
; PRIOR FILING DATE: 2000-11-27
; PRIOR FILING DATE: 2000-12-22
; PRIOR FILING DATE: 2000-12-22
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3936
; LENGTH: 1368
; TYPE: DNA
; ORGANISM: Enterococcus faecalis
US-09-815-242-3936
```

## Alignment Scores:

```
Pred. No.: 1.36e-51 Length: 1368
Score: 519.00 Matches: 152
Percent Similarity: 51.09% Conservative: 83
Best Local Similarity: 33.04% Mismatches: 181
Query Match: 22.96% Indels: 44
DB: 9 Gaps: 16
```

US-09-701-229-2 (1-448) x US-09-815-242-3936 (1-1368)

```
Qy 11 lLeValValGlyLeuGlyLysSerGlyMetSerLeuValArgTyrLeuAlaArgGly 30
Db 37 TTGGTTTAGCATAGTAAAGTGGTTCAGCCAGCAAACTCTTACATGATGATTAGT 96
Qy 31 LeuProPheAlaValAlaAspThrArg-----GluAsnPro-----ProGluLeu 45
Db 97 GCGCTCGTTACGTTAATGACGCAAAACAATTTGATCAAAACCTGACGCCAAGATT 156
Qy 46 AlaThrLeuArgAlaGlnTyrProGlnValGluValArgCysGlyGluLeuAspAlaGlu 65
Db 157 TTAACCTTG-----GGTATTCTGTTTACAGGGGGGCGCATCCAATTGAA 201
Qy 66 PheLeuCysSerAlaArgGluLeuTyrVal---SerProGlyLeuSerLeuArgThrPro 84
Db 202 TTGTTGATGAGATTTGAACTAATCGTTAAATAATCTGGTATCTTATCAAAATCCA 261
Qy 85 AlaLeuValGlnAlaAlaLysGlyValArgIleSerGlyAspIleAspLeuPheAla 104
Db 262 CTGTGGCAGAACGACCTAACTCGGAAATTCCTATCATCACTAGGTGGAATTAGCAGT 321
Qy 105 ArgGluAlaLysAlaProIleValAlaIleThrGlySerAsnAlaLysSerThrValThr 124
Db 322 CAAATTCGCAATGTCCAAATGTCGGCATTCAGGGCACCATAATGCAAAACCAACGACC 381
Qy 125 ThrLeuValGlyGluMetAlaValAlaAlaAspLys-----ArgValAlaValGly 141
Db 382 ACNATGATTGTT---TTACTGCTAAACGCTGACAGACGCGTGGTAGGACGTTGGCG 438
Qy 142 GlyAsnLeuGlyThrProAlaLeuAspLeu-----LeuAlaAspIle 156
Db 439 GGAATATTTGTTTCCAGCGAGTACGTCGGTGGCTCAAGAAGCAACGCGCATGATCTT 498
Qy 157 GluLeuTyrValLeuGluLeuSerSerPheGlnLeuGluThrCysAspArgLeuAsnAla 176
Db 157 GluLeuTyrValLeuGluLeuSerSerPheGlnLeuGluThrCysAspArgLeuAsnAla 176
```

```

Db 499 -----GTGATGGAACCTTCTAGCTTTTCACTTAATGGGAATGAGACGTTTCCACCCA 549
QY 177 GluValAlaThrValLeuAsnValSerGluAspHisMetAspArgTyrAspGlyMetAla 196
Db 550 CAAATGTCAGTAATTAACAATATTTTGAACACACTTGGATTATCATGTTCCGGGAAA 609
QY 197 AspTyrHisLeuAlaLysHisArgIlePheArgGlyAlaArg-----GlnValVal 213
Db 610 GAATATGTTGCTGCAAAATGGGCCATTCAAAAAACAATGACCGGTAGGACACCTTGATT 669
QY 214 ValAsn-----ArgAlaAspAlaLeuThrArgProLeuIleAlaAspThrVal 229
Db 670 TTAATGTAACAGTAGAGCTTCAACAGTTACGCCAAACACACAGCTCCCAACGTATGG 729
QY 230 ProCysTyrSerPheGlyLeuAsnLysProAspPheLysAlaPheGlyLeuIleGluGlu 249
Db 730 CCT-----TTTTCACGAAGAAGCAGTAGAAGGGGCTTATCTTTTA----- 771
QY 250 AspGlyGlnTyrTrpLeuAlaPheGlnPheAspLysLeuLeuProValGlyGluLeuLys 269
Db 772 GATGGGAAA-----TTATATTTTATGAAGAATATATATTCGCCCGCATGAGCTAGG 825
QY 270 IleArgGlyAlaHisAsnTyrSerAsnAlaLeuAlaAlaLeuAlaLeuGlyHisAlaVal 289
Db 826 ATTCTGTGTAGTACAAATATGAAATGCACTCGACGAGATTGTGTAGCTAAATTAAGA 885
QY 290 GlyLeuProPheAspAlaMetLeuGlyAlaLeuLysAlaPheSerGlyLeuAlaHisArg 309
Db 886 AATGTATCAATGTTTCAGATTAGACAAACTTTGAAAAACTTTTCGGGGCTTCCCATCGA 945
QY 310 CysGlnTyrValArgGlnGlyValSerTyrAspSerLysAlaThrAsn 329
Db 946 AGCAATATGTTGGCGAAGCTTCAGCAAGACGTTTTTATACGATTCAAAACCAACCAAT 1005
QY 330 ValGlyAlaAlaLeuAlaAlaIleGluGlyLeuGlyAlaAspIleAspGlyLysLeuVal 349
Db 1006 ATTTAGCTACAGAGTGGCTTAAGTGGTTGAC-----AACCAAGACTACTT 1056
QY 350 LeuLeuAlaGlyGlyAspGlyLysGlyAlaAspPheHisAspLeuArgGluProValAla 369
Db 1057 TTACTTGGCGGTGGCTTGACCGCGGTACTCATTTGATGAATTG---GTTCTGCTCTG 1113
QY 370 ArgPheCysArgAlaValValLeuLeuGlyArgAspAlaGlyLeuIleAlaGlnAlaLeu 389
Db 1114 CTGGGACTCAAGCAATGTTTATTTGAGAAACCAAGAAATAATGGCGGAAGCTGCT 1173
QY 390 -----GlyAsnAlaValProLeuValArgValAlaThrLeuAspGluAlaValArgGln 407
Db 1174 AAAAAAGCGAACATTTGAACAATTTTATTTGCTGAAATGTTCAACGCGGTACCATT 1233
QY 408 AlaAlaGluLeuAlaArgGlyAspAlaValLeuLeuSerProAlaCysAlaSerLeu 427
Db 1234 GCCTTTGATTATTCGAAAAAGATGATACATATTTACTATCATCCTGCTTGCAGAGTTGG 1293
QY 428 AspMetPheLysAsnPheGluGluArgGlyArgLeuPheAlaLysAlaValGluGluLeu 447
Db 1294 GACCAATACCGAATTTTGAAGTACGCGGGGAAGCCTTTATGCAAGCTGTTCAACAATTA 1353

```

RESULT 15

US-09-815-242-6539  
 ; Sequence 6539, Application US/09815242  
 ; Patent No. US20020061569A1

GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert  
 ; APPLICANT: Ohlson, Kari L.  
 ; APPLICANT: Zyskind, Judith W.  
 ; APPLICANT: Wall, Daniel  
 ; APPLICANT: Trawick, John D.  
 ; APPLICANT: Carr, Grant J.  
 ; APPLICANT: Yamamoto, Robert T.  
 ; APPLICANT: Xu, H. Howard  
 ; TITLE OF INVENTION: Identification of Essential Genes in  
 ; TITLE OF INVENTION: Prokaryotes

```

; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-03-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6539
; LENGTH: 1371
; TYPE: DNA
; ORGANISM: Enterococcus faecalis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1371)
; US-09-815-242-6539

```

```

Alignment Scores:
Pred. No.: 1,368-51 Length: 1371
Score: 519.00 Matches: 152
Percent Similarity: 51.09% Conservative: 83
Best Local Similarity: 33.04% Mismatches: 181
Query Match: 22.96% Indels: 44
DB: 9 Gaps: 16

```

US-09-701-229-2 (1-448) x US-09-815-242-6539 (1-1371)

```

QY 11 lleValValGlyLeuGlyLysSerGlyMetSerLeuValArgTyrLeuAlaArgGly 30
Db 37 TTGGTTTTAGGACTAGCTAAAGTGGTGTACGGCAGCGAAACTCTTACATGAGTTAGT 96
QY 31 LeuProPheAlaValAlaAspThrArg-----GluAsnPro-----ProGluLeu 45
Db 97 GCGCTCGTTACGTTAATGACGCAAAACAATTTGATCAAAACCCCTGACGCCCAAGATT 156
QY 46 AlaThrLeuArgAlaGlnTyrProGlnValGluValArgCysGlyGluLeuAlaGlu 65
Db 157 TTAACCTTG-----GGTATTCGTGTTGTACAGGGGGGCATCCATTGAA 201
QY 66 PheLeuCysSerAlaArgGluLeuTyrVal----SerProGlyLeuSerLeuArgThrPro 84
Db 202 TTGTTGGATGAAGAATTTGAACATAATCGTTAAAAATCTGGTATTCCTTATACAAATCCA 261
QY 85 AlaLeuValGlnAlaAlaLysGlyValArgIleSerGlyAspIleAspLeuPheAla 104
Db 262 CTGTGGCAGAGCACTAACTCGGAAAAATTCCTATCATCAACTGAGGTGGAATTAGCAGGT 321
QY 105 ArgGluAlaLysAlaProIleValAlaIleThrGlySerAsnAlaLysSerThrValThr 124
Db 322 CAATTTGCCGAATGTCCCAATGTCTGGCATTCGGGCACCAATGGCAAAACAACACGACCC 381
QY 125 ThrLeuValGlyGluMetAlaValAlaAlaAspLys-----ArgValAlaValGly 141
Db 382 ACGATGATTGGT---TTACTGCTAAACGCTGACAGACGGCTGGTGAGCAGCGTTTGGCG 438
QY 142 GlyAsnLeuGlyThrProAlaLeuAspLeu-----LeuAlaAspIle 156
Db 439 GGAATATTTGTTTCCAGCGAGTAGTGGTGGCTCAAGAAGCAACGCGCAAGGATGATCTT 498
QY 157 GluLeuTyrValLeuGluLeuSerPheGlnLeuGluThrCysAspArgLeuAsnAla 176
Db 499 -----GTGATGGAACCTTCTAGTTTTCAGTTAATGGGAATTTGAGACGTTTCCACCA 549

```

Search completed: August 14, 2003, 11:04:37  
Job time : 6736 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 14, 2003, 07:04:21 ; Search time 2292 Seconds  
(without alignments)  
4750.613 Million cell updates/sec

Title: US-09-701-229-2  
Perfect score: 2260  
Sequence: 1 MSIIASHDFRIVVGLGKSGM.....MFKNFEEGRGLFAKAVEELA 448

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 22781392 seqs, 1215238056 residues

Total number of hits satisfying chosen parameters: 45546872

Minimum DB seq length: 25  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+p2n.model -DEV=xlp  
-Q=/cgn2.1/USPTO.spool/US09701229/runat\_09082003.161157\_4393/app.query.fasta\_1.647  
-DB=EST -QFWT=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=25 -MAXLEN=2000000000  
-USER=US09701229.ecgn.1.1.4237 -runat\_09082003.161157\_4393 -NCPU=6 -ICPU=3  
-NO.MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- EST: \*
- 1: em\_estba:\*
  - 2: em\_esthum:\*
  - 3: em\_estin:\*
  - 4: em\_estnu:\*
  - 5: em\_estov:\*
  - 6: em\_estpl:\*
  - 7: em\_estro:\*
  - 8: em\_hic:\*
  - 9: gb\_estl:\*
  - 10: gb\_est2:\*
  - 11: gb\_hic:\*
  - 12: gb\_est3:\*
  - 13: gb\_est4:\*
  - 14: gb\_est5:\*
  - 15: em\_estfun:\*
  - 16: em\_estom:\*
  - 17: em\_gss\_hum:\*
  - 18: em\_gss\_inv:\*
  - 19: em\_gss\_pln:\*
  - 20: em\_gss\_vrt:\*
  - 21: em\_gss\_fun:\*
  - 22: em\_gss\_mam:\*
  - 23: em\_gss\_mus:\*
  - 24: em\_gss\_pro:\*
  - 25: em\_gss\_rod:\*
  - 26: em\_gss\_phg:\*
  - 27: em\_gss\_vri:\*
  - 28: gb\_gss1:\*

29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
c 1	1084	48.0	892	29	BZ579116	BZ579116 msh2.6205
c 2	1081.5	47.9	1445	29	BZ568959	BZ568959 pacs2.164
c 3	603.5	26.7	1171	29	BZ559806	BZ559806 pacs2.164
c 4	391	17.3	452	10	BE184674	BE184674 PM4-HT068
c 5	313	13.8	928	29	BZ567557	BZ567557 pacs2.164
c 6	240	10.6	714	28	BH374901	BH374901 AG-ND-127
c 7	233	10.3	748	12	BJ172647	BJ172647 BJ172647
c 8	231	10.2	760	28	BH367641	BH367641 AG-ND-127
c 9	226	10.0	1194	29	BZ550283	BZ550283 pacs1-60
c 10	213.5	9.4	572	28	BH795426	BH795426 BMBAC376G
c 11	210	9.3	965	29	CNS070MR	AL423817 T3 end of
c 12	187	8.3	595	29	CNS070MR	AL423816 T7 end of
c 13	186	8.2	616	28	BH383040	BH383040 AG-ND-127
c 14	172.5	7.6	495	28	BH369957	BH369957 AG-ND-138
c 15	172	7.6	378	28	BH382726	BH382726 AG-ND-127
c 16	160.5	7.1	651	13	BQ504834	BQ504834 EST612249
c 17	136.5	6.0	689	10	BG592398	BG592398 EST491076
c 18	136	6.0	684	28	AQ968325	AQ968325 LFRJB56TR
c 19	132.5	5.9	1214	29	BZ572325	BZ572325 msh2.2585
c 20	129	5.7	422	29	BZ406343	BZ406343 OGAAC90TF
c 21	125.5	5.6	838	29	CC115454	CC115454 NDL.42F3
c 22	124.5	5.5	794	14	CB853399	CB853399 UT-CF-FNO
c 23	124.5	5.5	930	29	BZ675115	BZ675115 PUBDE10TD
c 24	123	5.4	730	29	CC063642	CC063642 fgma003d0
c 25	122.5	5.4	146	9	AL719411	AL719411 AL719411
c 26	118	5.2	699	29	CC063650	CC063650 fgma003d0
c 27	117.5	5.2	685	29	CC140248	CC140248 NDL.42F4
c 28	114.5	5.1	191	14	CA548165	CA548165 CO802A11
c 29	112.5	5.0	817	29	CNS0110M	AL146223 Anopheles
c 30	110	4.9	1049	10	BE216920	BE216920 EST0463 T
c 31	106.5	4.7	584	10	BE760975	BE760975 SNQvL2CAS
c 32	104.5	4.6	673	12	BI718971	BI718971 1031033F0
c 33	103.5	4.6	884	29	BZ564204	BZ564204 pacs2.164
c 34	103.5	4.6	1108	29	BZ559252	BZ559252 pacs2.164
c 35	103.5	4.6	1482	29	BZ557157	BZ557157 pacs1-60
c 36	102.5	4.5	1522	12	BM320865	BM320865 rockefell
c 37	102	4.5	850	14	CB650740	CB650740 OSINEB15F
c 38	101.5	4.5	597	9	AW200691	AW200691 se92c10.y
c 39	101.5	4.5	876	10	BE216958	BE216958 EST0501 T
c 40	101.5	4.5	1031	29	BZ563054	BZ563054 pacs2.164
c 41	101	4.5	761	12	BM076245	BM076245 TrEST-A11
c 42	101	4.5	775	10	BG368290	BG368290 HVSME1001
c 43	101	4.5	829	10	BF264916	BF264916 HV-Cra001
c 44	101	4.5	1118	29	BZ548884	BZ548884 pacs1-60
c 45	100	4.4	822	11	CNS09MEH	EX065109 Single re

ALIGNMENTS

RESULT 1

BZ579116/c

LOCUS

DEFINITION

ACCESSTON

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

1 (bases 1 to 892)

msh2.6205.y2 msh Pseudomonas aeruginosa genomic clone msh2.6205,  
genomic survey sequence.

BZ579116 892 bp DNA linear GSS 17-DEC-2002

BZ579116.1 GI:27214177

Pseudomonas aeruginosa

Pseudomonas aeruginosa

Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;

Pseudomonadaceae; Pseudomonas.

1 (bases 1 to 892)

**AUTHORS** Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M., Burns,J.L., Kaul,R. and Olsen,M.V.  
**TITLE** Whole-Genome-Sequence variation among multiple isolates of *Pseudomonas aeruginosa* library  
**JOURNAL** J. Bacteriol., (2002) In press  
**COMMENT** Contact: Chris K. Raymond  
 Genome Center  
 University of Washington  
 Box 352145, Seattle, WA 98105-2145, USA  
 Tel: 2062216954  
 Fax: 2066857244  
 Email: craymond@u.washington.edu  
 Class: shotgun.

# FEATURES

source  
 1. -892  
 Location/Qualifiers  
 /organism="Pseudomonas aeruginosa"  
 /mol\_type="genomic DNA"  
 /strain="M5H"  
 /db\_xref="taxon:287"  
 /clone="msh2\_6205"  
 /clone\_lib="msh"  
 /note="Environmental isolate. Whole genomic shotgun library."  
 BASE COUNT 173 a 277 c 276 g 160 t 6 others  
 ORIGIN

Alignment Scores:  
 Pred. No.: 3,73e-112 Length: 892  
 Score: 1084.00 Matches: 230  
 Percent Similarity: 91.37% Conservative: 3  
 Best Local Similarity: 90.20% Mismatches: 19  
 Query Match: 47.96% Indels: 5  
 DB: 29 Gaps: 2

US-09-701-229-2 (1-448) x B2579116 (1-892)

QY 23 ValArgTyrLeuAlaArgArgGlyLeuProPheAlaValValAspThrArgGluAsnPro 42  
 DB 839 GTGCGTTACTGG---CGCGNGGGNTGCTTTTGGCGTGGTCAATACC---GAAAGGANC 786  
 QY 43 ProGlu-LeuAlaThrLeuArgAlaGlnTyrProGlnValGluAlaArgCysGlyGluLe 62  
 DB 785 CCGCAAGTGGCCACCTTGGTCCCGCATATCG-CAAGTTGAAAGTGGTGGCGGCAACT 727  
 QY 62 uAspAlaGluPheLeuCysSerAlaArgGluLeuTyrValSerProGlyLeuSerLeuAr 82  
 DB 726 CAACGCGAGTTCTTTTGTCTGCGCGCGCAACTTTACGTGAC-CCCGCGTTCGTGTCG 668  
 QY 82 gThrProAlaValGlnAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 102  
 DB 667 CACCNCGGCTGTGTACAGCCCGCGCGGAAAGCGCGCATCTCCGTGACATCGATCT 608  
 QY 102 uPheAlaArgGluAlaLysAlaProIleValAlaIleThrGlySerAsnAlaLysSerTh 122  
 DB 607 CTTCCGCGGAGGCGAGGCGAGGCGGATCGTCCCATACCGGTTCCACGCGGAAGAC 548  
 QY 122 rValThrThrLeuValGlyGluMetAlaValAlaAlaAspLysArgValAlaValGlyCl 142  
 DB 547 CGTGACCACTCTGTGTGGGAAATGGCGTGGCGCGGCAAGCGTGTGCGCGTGGCGGG 488  
 QY 142 yAsnLeuGlyThrProAlaLeuAspLeuLeuAlaAspAlleGluLeuTyrValLeuGl 162  
 DB 487 CAACTCGGACCCCGCGCTGACCTGTGGCGGACGACATCGAGCTGTACGTTGGA 428  
 QY 162 uLeuSerSerPheGlnLeuGluThrCysAspArgLeuAsnAlaGluValAlaIleThrVal 182  
 DB 427 GCTGTGAGCTTCCAGCTGGAACCTGCGATCGCTCAACCGGAGGTGGCGACCTGCT 368  
 QY 182 uAsnValSerGluAspHisMetAspArgTyrAspGlyMetAlaAspTyrHisLeuAlaLy 202  
 DB 367 GAACGTGCGGAAGACCATATGATGATGCTAGCAGCGCATGGTGTGACTACCACTGGCCA 308  
 QY 202 sHisArgIlePheArgGlyAlaArgGlnValValValAsnArgAlaAspAlaLeuThrAr 222

DB 307 GCACCGGATCTTCCGCGTCCCGCAGCGTCTGGTGAATCGCGCGATCGCTGACCGG 248  
 QY 222 gProLeuIleAlaAspThrValProCysTrpSerPheGlyLeuAsnLysProAspPheLy 242  
 DB 247 ACCGCTGATCGCGCATACCGTCCGTCGTGGTGGCTGCGCCCTGACAAAGCCGACTTCAA 188  
 QY 242 sAlaPheGlyLeuIleGluGluAspGlyGlnLysTrpLeuAlaPheGlnPheAspLysLe 262  
 DB 187 GCGTTTGGCCTGATCGAGGAGACAGCGGCGAGAGTGGTGGTGGCTTCCAGTTCCACAAGCT 128  
 QY 262 uLeuProValGlyGluLeuLysIleArgGlyAlaHisAsnTyr 276  
 DB 127 GCTGCCGTTGGCGAACTGAAGATCGTGGCGCCCAACTAT 85

RESULT 2  
 B2568959/c  
 LOCUS  
 DEFINITION pacs2-164\_8171.y2 pacs2-164 Pseudomonas aeruginosa genomic clone  
 pacs2-164\_8171, genomic survey sequence.  
 ACCESSION B2568959  
 VERSION B2568959.1 GI:27202799  
 KEYWORDS GSS.  
 SOURCE Pseudomonas aeruginosa  
 ORGANISM Pseudomonas aeruginosa  
 Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
 Pseudomonadaceae; Pseudomonas.

REFERENCE 1 (bases 1 to 1445)  
 AUTHORS Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M., Burns,J.L., Kaul,R. and Olsen,M.V.  
 TITLE Whole-Genome-Sequence variation among multiple isolates of *Pseudomonas aeruginosa* library  
 JOURNAL J. Bacteriol., (2002) In press  
 COMMENT Contact: Chris K. Raymond  
 Genome Center  
 University of Washington  
 Box 352145, Seattle, WA 98105-2145, USA  
 Tel: 2062216954  
 Fax: 2066857244  
 Email: craymond@u.washington.edu  
 Class: shotgun.

FEATURES  
 source  
 1. 1445  
 Location/Qualifiers  
 /organism="Pseudomonas aeruginosa"  
 /mol\_type="genomic DNA"  
 /strain="2-164"  
 /db\_xref="taxon:287"  
 /clone="pacs2-164\_8171"  
 /note="clinical isolate 2-164 Whole genomic shotgun library."  
 BASE COUNT 308 a 424 c 412 g 293 t 8 others  
 ORIGIN

Alignment Scores:  
 Pred. No.: 1.49e-111 Length: 1445  
 Score: 1081.50 Matches: 241  
 Percent Similarity: 87.63% Conservative: 7  
 Best Local Similarity: 85.16% Mismatches: 25  
 Query Match: 47.85% Indels: 15  
 DB: 29 Gaps: 4  
 US-09-701-229-2 (1-448) x B2568959 (1-1445)

QY 69 SerAlaArgGluLeuTyrVal-----SerProGlyLeuSerLeuArgThr 83  
 DB 904 AACTCACCAGAGCTTTTGTGGCGGCAACTTTACTAGCCCGGCTTCTGTTGTC-ACC 846  
 QY 84 ProAlaLeuValGlnAlaAlaLysGlyValArgIleSerGlyAspIleAspLeuPhe 103  
 DB 845 CCGCGCTGTACAGCGCGGCAAGGCTGGCATCTCGTAC-----ATCGATCTT--- 798  
 QY 104 AlaArgGluAlaLysAlaProIleValAlaIleThrGlySerAsnAlaLysSerThrVal 123

```

Db 797 TCAGCGCAGGCGAAGGCCCGA---TCGTNGCATCACGGTTCCAAACGGCGAAGAGCACCGTG 741
Qy 124 ThrThrLeuValGlyGluMetAlaValAlaAspLysArgValAlaValAlaGlyGlyAsn 143
Db 740 ACAACCTT-GTGGCGGAATGACGGTGGC-GCGNACAGCGGTGCGGCTC-GNCGGCAAC 684
Qy 144 LeuGlyThrProAlaLeuAspLeuLeuAlaAspAspIleGluLeuThrValLeuGluLeu 163
Db 683 CTCGGCAC-CCGGCGCTCGACCTGCTGCCGACGACATCGAGCTGTACGTGTGGAGCTG 625
Qy 164 SerSerPheGlnLeuGluThrCysAspArgLeuAsnAlaGluValAlaThrValLeuAsn 183
Db 624 TCAGAGCTCCAGCTGGAACCTCGCATCGCTCAACGCCGAGGTGGCGACCGTCTGAAC 565
Qy 184 ValSerGluAspHisMetAspArgTyrAspGlyMetAlaAspTyrHisLeuAlaLysHis 203
Db 564 GTCAGCGAAGACCATATGATGCTGCTACGACGGCATGGCTGACTACACCTGGCCAGCAC 505
Qy 204 ArgIlePheArgGlyAlaArgGlnValValValAsnArgAlaAspAlaLeuThrArgPro 223
Db 504 CGGATCTTCGCGGTGCCGCGCAGGTGCTGTGAATCGCGCGATGCCCTGACCCGACCG 445
Qy 224 LeuIleAlaAspThrValProCysTrpSerPheGlyLeuAsnLysProAspPheLysAla 243
Db 444 CTGATCGCGGATACCGTGGCGGTGCTGTGCTGCTGCGCTGAACAGCGGACTTCAAGGCT 385
Qy 244 PheGlyLeuIleGluAspGlyGlnLysTrpLeuAlaPheGlnPheAspLysLeuLeu 263
Db 384 TTCGCGCTGATCGAGGAGACGGCCAGAGTGGTGGCTTCCAGTTCGACAGCTGCTG 325
Qy 264 ProValGlyGluLeuLysIleArgGlyAlaHisAsnTyrSerAsnAlaLeuAlaAlaLeu 283
Db 324 CCGGTTCGGCAACTGAAGATCCGTGGCGCCACAACTATTCCAAACGCGCTCGCGCGCTG 265
Qy 284 AlaLeuGlyHisAlaValGlyLeuProPheAspAlaMetLeuGlyAlaLeuLysAlaPhe 303
Db 264 GCGTGGGCGCATCGGTGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 205
Qy 304 SerGlyLeuAlaHisArgCysGlnTrpValArgGluArgGlnGlyValSerTyrTyrAsp 323
Db 204 TCCGCGCTGGCTCATCGCTGCCAGTGGGTACGCGAGCGGCGGCGGCTGAGCTACTAGAC 145
Qy 324 AspSerLysAlaThrAsnValGlyAlaAlaLeuAlaAlaIleGluGlyLeuGlyAlaAsp 343
Db 144 GATTCGAAGGCCAACCAACGTGCGGCGCGCTGCGCGCGATCGAGGGGCTGGGTGCGGAC 85
Qy 344 IleAspGly 346
Db 84 ATCGACGGG 76

```

## RESULT 3

BZ559806 1171 bp DNA linear GSS 17-DEC-2002  
 LOCUS pacs2-164\_1863.x1 pacs2-164 Pseudomonas aeruginosa genomic clone  
 DEFINITION pacs2-164\_1863, genomic survey sequence.

## ACCESSION

BZ559806

## VERSION

BZ559806.1 GI:27176698

## KEYWORDS

GSS.

## SOURCE

Pseudomonas aeruginosa

## ORGANISM

Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;

1 (bases 1 to 1171)

Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,

Burns, J.L., Kaul, R. and Olsen, M.V.

Whole-Genome-Sequence variation among multiple isolates of

Pseudomonas aeruginosa library

J. Bacteriol., (2002) In press

Contact: Chris K. Raymond

Genome Center

University of Washington

Box 352145, Seattle, WA 98105-2145, USA

Tel: 2062216954

Fax: 2066857244

Email: craymond@u.washington.edu

Class: shotgun

## FEATURES

## source

1. 1171

Location/Qualifiers

/organism="Pseudomonas aeruginosa"

/mol\_type="genomic DNA"

/strain="2-164"

/db\_xref="taxon:287"

/clone="pacs2-164\_1863"

/clone\_lib="pacs2-164"

/note="clinical isolate 2-164 Whole genomic shotgun library."

BASE COUNT 218 a 325 c 329 g 298 t 1 others

## ORIGIN

## Alignment Scores:

Pred. NO.: 1.64e-57 Length: 1171

Score: 603.50 Matches: 132

Percent Similarity: 83.13% Conservative: 6

Best Local Similarity: 79.52% Mismatches: 25

Query Match: 26.70% Indels: 3

DB: 29 Gaps: 1

US-09-701-229-2 (1-448) x BZ559806 (1-1171)

Qy 276 TyrSerAsnAlaLeuAlaAlaLeuAlaLeuGlyHisAlaValGlyLeuProPheAspAla 295

Db 48 TATTCCAACGGCTCGCGGAGCTGGACATGGGCGCATGGGTCGCCCTCTTTACGCC 107

Qy 296 MetLeuGlyAlaLeuLysAlaPheSerGlyLeuAlaHisArgCysGlnTrpValArgGlu 315

Db 108 ATGCTCGCGCGCTGAAAGCGTCTTGGGAGCTGCTCATTCCTCCAGTGGGTACGCGAG 167

Qy 316 ArgGlnGlyValSerTyrTyrAspSerLysAlaThrAsnValGlyAlaAlaLeuAla 335

Db 168 CGGAGGCGGTGAGCTACTAGACGATTTCCAAAGCCACCAACGTCGGCGCCCTGGCG 227

Qy 336 AlaIleGluGlyLeuGlyAlaAspIleAspGlyLysLeuValLeuLeuGlyGlyLys 355

Db 228 GCGATCGAGGCGGTGGTGGCGCATCGACGGCAAGCTGCTGCTGCTGCTGCTGCTGCTG 287

Qy 356 GlyLysGlyAlaAspPheHisAspLeuArgGluProValAlaArgPheCysArgAlaVal 375

Db 288 GGCAAGGCGCGGATTTTCATGACCTCGCGAGCGGTCGCGCTCTCTCGCGCGGTG 347

Qy 376 ValLeuLeuGlyArgAspAlaGlyLeuIleAlaGlnAlaLeuGlyAsnAlaValProLeu 395

Db 348 GTACTGCTTGTGCTGACACCTAGCTGATTCACGAGGACCTGGGCAACGCGGACCGCTG 407

Qy 396 ValArgValAlaThrLeuAspGluAlaValArgGlnAlaAlaGluLeuAlaArgGluGly 415

Db 408 GTGCGCGTCAACGCTGGAGAAAGCATTTCCGAGGCGCGGAGCTGGCGCGCGAAGC 467

Qy 416 Asp-AlaValLeuLeuSerProAlaCysAla--SerLeuAspMetPhe-LysAsnPheG 434

Db 468 GATAATAGTCTGTGGGATTTCCCGGCCCTTCCCAACCCGCTGATGTTAAAAAACCTTT 527

Qy 434 luGluArgGlyArg 438

Db 528 TAAAAAAGCGCG 541

## RESULT 4

BE184674

## LOCUS

PM4-HT0688-050500-002-d10 HT0688

## DEFINITION

PM4-HT0688-050500-002-d10 HT0688 Homo sapiens cDNA, mRNA sequence.

## ACCESSION

BE184674

## VERSION

BE184674.1

## KEYWORDS

EST.

## SOURCE

Homo sapiens (human)

## ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;



```

Db      464  GGTCTGCGACTCAGCAGCGCGGTTTACAAAGCTTTTCGACCTGGTCCAGAAAGTCGTAC 405
Qy      252  lnLysTrpLeuAlaPheGlnPheAspLysLeuLeuProValGlyGluLeuLysIleArg 272
Db      404  AGAAGTGGCTGGCTAGCTGTTACGCAAAATAATTCGGCTGCTGATGTAAGATCCGTC 345
Qy      272  yAlaHisAsnTyrSerAsnAlaLeuAlaLeuAlaLeuGlyHisAlaValGlyLeu 292
Db      344  GAGCAGCGGACTATTATACGCTTCCGCGCTGGCGTGGCGCATCGCGTCGTGC 285
Qy      292  roPhe-AspAlaMetLeuGlyAlaLeuLysAlaPheSerGlyLeuAlaHisArgCysGln 311
Db      284  CGTTCTACAGCATATGATGGCGCACACAGGTCATCCAGCTGTTCTCTATCCTCGCCA 225
Qy      312  TrpValArgGluArgGlnGlyValSerTyrTyr-AspAspSerLysAlaThrAsnValG 331
Db      224  TGAGTACCCTAACGACAGGCTCTGAGTATGTTCCGACGATTCACGAGCCGCTCTGGT 165
Qy      331  yAlaAlaLeuAlaAlaIleGluGlyLeuGlyAlaAspIleAspGlyLysLeuValLe 351
Db      164  CACGCGCCTGTTTCGATCGAGTGGATGGTGTGGACATTTTCATCAACCTCATCTGTG 105
Qy      351  uAlaGlyClyAspGlyLysGlyAlaAspPheHisAspLeuArgGluProValAlaArg 370
Db      104  CGCAGACGACGACGACAAAGGG---GATCCACTAGTTCTAGAGCGCGCCGACCGGTG 48
Qy      371  -----PheCys 372
Db      47  GAGCTCCAGCTTTGT 32

RESULT 6
BH374901/c
LOCUS      714 bp      DNA      linear      GSS 10-DEC-2001
DEFINITION AG-ND-127J23-TR ND-TAM Anopheles gambiae genomic clone AG-ND-127J23
            , genomic survey sequence.
ACCESSION  BH374901
VERSION    BH374901.1 GI:17321043
KEYWORDS   GSS.
SOURCE     Anopheles gambiae (African malaria mosquito)
ORGANISM   Anopheles gambiae
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
            Anopheles.
REFERENCE  1 (bases 1 to 714)
            Hong,Y.S., Hogan,J.R., Wang,X., Sarkar,A., Sim,C., Loftus,B.J., Ren
            ,C., Huff,E.R., Carlile,J.L., Black,K., Zhang,H.-B., Gardner,M.J.
            and Collins,F.H.
            Construction of a BAC library and generation of BAC end
            sequence-tagged connectors for genome sequencing of the African
            malaria mosquito Anopheles gambiae
            Mol. Genet. Genomics 268 (6), 720-728 (2003)
            22542063
            PUBMED
COMMENT    Other GSSs: AG-ND-127J23.TF
            Contact: Brendan J Loftus
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0208
            Fax: 301 838 3543
            Email: b.loftus@tigr.org
            This clone is from an A. gambiae BAC library (ND-TAM) provided by
            F.H. Collins and sequenced by The Institute for Genomic Research
            (TIGR). The BAC library was generated from A. gambiae PEST strain
            DNA. All DNA was extracted from newly hatched first instar larvae
            to minimize the inclusion of DNA from microorganisms that inhabit
            the gut. The DNA is derived from mixed sexes of larvae. The BAC
            library was constructed at Texas A&M University BAC Center
            University, College Station, Texas 77843-2123, USA using a HindIII
            partial digest.
            Seg primer: M13 Rev
            Class: BAC ends.
            Location/Qualifiers

```

```

source
1. 714
/organism="Anopheles gambiae"
/mol_type="genomic DNA"
/strain="PEST"
/db_xref="taxon:7165"
/clone="AG-ND-127J23"
/clone_lib="ND-TAM"
/note="vector: pECBAC1; Site_1: HindIII"
BASE COUNT 212 a 140 c 106 g 256 t
ORIGIN
Alignment Scores:
Pred. NO.: 1.23e-16 Length: 714
Score: 240.00 Matches: 67
Percent Similarity: 46.56% Conservative: 48
Best Local Similarity: 27.13% Mismatches: 94
Query Match: 10.62% Indels: 38
DB: 28 Gaps: 7
US-09-701-229-2 (1-448) x BH374901 (1-714)
Qy      171  CysAspArgLeu-----AsnAlaGluValAlaThrValLeuAsnValSerGluAsp 187
Db      706  TGTGATAATATTCAGTATTTTCAGCCATATATCAGTTTCTCTAAATCTGTCTCAGAC 647
Qy      188  HisMetAspArgTyrAsp---GlyMetAlaAspTyrHisLeuAlaLysHisArgIlePhe 206
Db      646  CATTAGACCAGTATTAACAATATGAAGAATACGATTAAGCAAAATTCAGAAATTGCA 587
Qy      207  ArgGlyAlaArgGln-----ValValValAsnArgAlaAspAlaLeuThrArgPro 223
Db      586  GAAATCAGGAAAATGATAATTATTTTCATCTACAATAAAGATGATGAATCAGCAGAAA 527
Qy      224  LeuIle-----AlaAspThrValProCysTyrTrpSerPheGlyLeuAsn 237
Db      526  ATTCTTCAGTCTCTGGAGATTAAATGCTACAATGATTCCT----- 488
Qy      238  LysProAspPheLysAlaPheGlyLeuIleGlu----- 249
Db      487  -----TTCTCTATGAACAGAGAGCTAAGTGAAGTGGTTATTCAATT 446
Qy      250  AspGlyGlnLysTrpLeuAlaPheGlnPheAspLysLeuLeuProValGlyGluLeuLys 269
Db      445  GACAATGAATTAGTCATTAATACTTCAGGACGACTTCAGAATGAAATTTCTGATCTTCA 386
Qy      270  IleArgGlyAlaHisAsnTyrSerAsnAlaLeuAlaLeuAlaLeuGlyHisAlaVal 289
Db      385  CTTGTAGAAACCAATAATGTTGCCAATAGTTAGCTGCAAGTATAGCAGGTAACCTGCTA 326
Qy      290  GlyLeuProPheAspAlaMetLeuGlyAlaLeuLysAlaPheSerGlyLeuAlaHisArg 309
Db      325  AATATTAGTAATAAGATATCCGCAATAGTCTGATGACATTTACGGCTGTCGACAGA 266
Qy      310  CysGlnTrpValArgGluArgGlnGlyValSerTyrTyrAspAspSerLysAlaThrAsn 329
Db      265  TTGGACGAGGTGTCAGTTATTAAATCATGCTGAATACATCAATGACAGTAAGCAACCA 206
Qy      330  ValGlyAlaAlaLeuAlaAlaIleGluGlyLeuGlyAlaAspIleAspGlyLysLeuVal 349
Db      205  GTAATGCTGATATATGCACTGAA-----AGTGTGAATATCCTGCAATA 158
Qy      350  LeuLeuAlaGlyGlyAspGlyLysAlaAspPheHisAspLeuArgGluProValAla 369
Db      157  TGGATTGTTGCTGGAATAGATAAGAAATCAGTATACAGAAATTCAGGATCTGTGTAAG 98
Qy      370  ArgPheCysArgAlaValValLeuLeuGlyArgAspAlaGlyLeuIleAlaGlnAlaLeu 389
Db      97  AAGAAAGTTCGTGCGATGCTGGATGGATGATGAAGAAAGATCATTCAGTCTCTC 38
Qy      390  GlyAsnAlaValProLeuVal 396
Db      37  AGAAACAAAAGGATTGATT 17

```

FEATURES

```

RESULT 7
BJ172647/c
LOCUS
DEFINITION
    BJ172647 748 bp mRNA linear EST 24-JAN-2002
    gametophores Physcomitrella patens subsp. patens cDNA clone
    pph33b18 3', mRNA sequence.
ACCESSION
VERSION
SOURCE
ORGANISM
    Physcomitrella patens subsp. patens
    Eukaryota; Viridiplantae; Streptophyta; Bryophyta;
    Bryopsida; Funariidae; Funariales; Funariaceae; Physcomitrella.
    1 (bases 1 to 748)
REFERENCE
AUTHORS
    Fujita,T., Shin-I,T., Seki,M., Kamiya,A., Uchiyama,I., Nishiyama,T.
    , Carninci,P., Hayashizaki,Y., Shinozaki,K., Kohara,Y. and Hasebe
    ,M.
TITLE
    Comparison of the moss Physcomitrella patens genome with flowering
    plants genome
JOURNAL
COMMENT
    Contact: Tadasu Shin-I
    Center For Genetic Resource Information
    National Institute of Genetics
    1111 Yata, Mishima, Shizuoka 411-8540, Japan
    Tel: 81-559-81-6856
    Fax: 81-559-81-6855
    Email: tshin@genes.nig.ac.jp
    A backbone of the vector is basically from pBluescript(KS), that
    was in vivo excised from a modified lps phage vector (Mo bi Tec,
    Germany). 5' end of the cDNA that was digested with XhoI was
    ligated to SalI site of the vector and the 3' end including polyA
    tail was ligated to BamHI site of the vector. cDNA insert could be
    amplified with conventional T7 and T3 primers. This full-length
    cDNA library was generated basically according to the method
    described in The Plant J 15, 707-720 (1998) Seki M. et al.
    Protonemata were blended by the POLYTRON, and then cultivated on
    the BCDATG medium for 13-14 days under the continuous light.

FEATURES
    source
        1..748
        /organism="Physcomitrella patens subsp. patens"
        /mol_type="mRNA"
        /sub_species="patens"
        /db_xref="taxon:145481"
        /clone="pph33b18"
        /tissue_type="mixture of chloronemata and young
        gametophores with 2 to 5 leaves"
        /clone_lib="full length cDNA library, chloronemata and
        young gametophores"
BASE COUNT
    214 a 192 c 144 g 198 t

Alignment Scores:
Pred. No.:      8.19e-16      Length:      748
Score:          233.00      Matches:      68
Percent Similarity: 51.46%      Conservative: 20
Best Local Similarity: 39.77%      Mismatches: 69
Query Match:    10.31%      Indels:      14
DB:              12      Gaps:          5

US-09-701-229-2 (1-448) x BJ172647 (1-748)

QY 281 AlaAlaLeuAlaGlyHisAlaValGlyLeuPropheAspAlaMetLeuGlyAlaLeu 300
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 735 GCTCTTCTTGCCATCTCTTAGATGTCAGTCAGTACAGAGATATTCAGCGGCTATT 676

QY 301 LysAlaPheSerGlyLeuAlaHisArgCysGlnTrpVal---ArgGluArgGlnGlyVal 319
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 675 CCCCACTTAAGGCTCTCCACATCGTATGGAATTGTGCATCAAGACGACCAAGGGGTG 616

QY 320 SerTyrTrpAspSerLysAlaThrAsnValGlyAlaAlaLeuAlaAlaTleGluGly 339
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 615 GTGTGGGTGAATGATAGCAAGCCACGATGTGGATTCCACTTCAGTTGGCATAGAAGC 556

```

```

QY 340 LeuGlyAlaAspIleaspGlyLysLeuValLeuLeuAlaGlyGlyAspGlyLys----- 357
    ::||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 555 ATT-----GTGGGAAAGAAAGACAGTGGTGTGTAGGGGCCCAAGCAAGGTGGA 505

QY 358 -----GlyAlaAspPheHisAspLeuArgGluProValAlaAlaArgPheCysArg 373
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 504 GTAGGAAGCGGAGGACTTGGATTGGTGTCTAGTTGAATCCCTACATCTCAT---CGG 448

QY 374 AlaValValLeuLeuGlyArgAspAlaGlyLeuAlaGlnAlaLeu-----Gly 390
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 447 GCTGTGTACTTTTGGAGCATCGGCCCTTGCATTGAACAGAGTTGGTCAAGCAGGT 388

QY 391 AsnAlaValProLeuValArgValAlaThrLeuAspGluAlaValArgGlnAlaGlu 410
    ::||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 387 TTGATGATCATCATGCATATACAGAACTTTTTCAGATGCTGCTGCTGCTGCTGCTCT 328

QY 411 LeuAlaArgGluGlyAspAlaValLeuLeuSerProAlaCysAlaSerLeuAspMetPhe 430
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 327 TTGGCACAACTGGAGATGCTGTCTCTGAGTCTGCTGCTGCTGCTGCTGCTGCTGCT 268

QY 431 LysAsnPheGluGluArgGlyArgLeuPheAla 441
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 267 AAAAATTTTGAGCATCGTGGCGAAGTTTTCGCC 235

RESULT 8
BH367641/c
LOCUS
DEFINITION
    AG-ND-127N18.TR ND-TAM Anopheles gambiae genomic clone AG-ND-127N18
    , genomic survey sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
    BH367641.1 GI:17313685
    GSS.
ORGANISM
    Anopheles gambiae (African malaria mosquito)
    Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
    Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
    Anopheles.
REFERENCE
    1 (bases 1 to 760)
    Hong,Y.S., Hogan,J.R., Wang,X., Sarkar,A., Slin,C., Loftus,B.J., Ren
    ,C., Huff,E.R., Carlile,J.L., Black,K., Zhang,H.-B., Gardner,M.J.
    , and Collins,F.H.
    Construction of a BAC library and generation of BAC end
    sequence-tagged connectors for genome sequencing of the African
    malaria mosquito Anopheles gambiae
    Mol. Genet. Genomics 268 (6), 720-728 (2003)
    22542063
    PUBMED
    12655398
    Other_GSSs: AG-ND-127N18.TF
    Contact: Brendan J Loftus
    Department of Eukaryotic Genomics
    The Institute for Genomic Research
    9712 Medical Center Dr., Rockville, MD 20850, USA
    Tel: 301 838 0208
    Fax: 301 838 3543
    Email: b.loftus@tigr.org
    This clone is from an A. gambiae BAC library (ND-TAM) provided by
    F.H. Collins and sequenced by The Institute for Genomic Research
    (TIGR). The BAC library was generated from A. gambiae PEST strain
    DNA. All DNA was extracted from newly hatched first instar larvae
    to minimize the inclusion of DNA from microorganisms that inhabit
    the gut. The DNA is derived from mixed sexes of larvae. The BAC
    library was constructed at Texas A&M University BAC Center
    University, College Station, Texas 77843-2123, USA using a HindIII
    partial digest.
    Seq primer: M13 Rev
    Class: BAC ends.
    Location/Qualifiers
        1..760
        /organism="Anopheles gambiae"
        /mol_type="genomic DNA"
        /strain="PEST"
        /db_xref="taxon:7165"
        /clone="AG-ND-127N18"

```

/clone\_lib="ND-TAM"  
/note="vector: pECBAC1; site\_1: HindIII"  
BASE COUNT 222 a 155 c 112 g 271 t  
ORIGIN

## Alignment Scores:

Pred. No.: 1.41e-15 Length: 760  
Score: 231.00 Matches: 70  
Percent Similarity: 47.10% Conservative: 52  
Best Local Similarity: 27.03% Mismatches: 100  
Query Match: 10.22% Indels: 39  
DB: 28 Gaps: 7

US-09-701-229-2 (1-448) x BH367641 (1-760)

```

QY 167 GlnLeuGluThrCysAspArgLeuAsnAlaGluValAlaThrValLeuAsnValSerGlu 186
    |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 759 CAGTTGGATGATACAGAATTCAGACCATATATAGTTGCTTCTAAATCTCTCTCAG 700
    |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 187 AspHisMetAspArgTyrAsp---GlyMetAlaAspTyrHisLeuAlaLysHisArgIle 205
    |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 699 GACCATTTAGACCAGTATACAAATATATCAAGATACGATAGCAAAATTCAGAAAT 640
    |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 206 PheArgGlyAlaArgGln-----ValValValAsnArgAlaAspAlaLeuThrArg 222
    |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 639 GCAGAAATCAGGAAATGATAATATTTCATCTACATAAAGATGATGAATGAGCCAG 580
    |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 223 ProLeuIle-----AlaAspThrValProCysTyrSerPheGlyLeu 236
    |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 579 AAATCTCTAGCTCTCGGAGATTAAATGCTACATGATTCCT----- 538
    |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 237 AsnLysProAspPheLysAlaPheGlyLeuIleGluLysAspGlyGlnLysTrp----- 254
    |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 537 -----TTCTCTATGAAAGAGAGCTAAGTGA-AGTGTGTATTC 500
    |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 255 -----LeuAlaPheGlnPheAspLysLeuLeuProValGlyGlu 267
    |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 499 AATTGACCATGATTAAGTCTAATTAACCTCAGGAC-GACTTTCAGAAATTTCTGAT 441
    |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 268 LeuLysIleArgGlyAlaHisAsnTyrSerAsnAlaLeuAlaAlaLeuGlyHis-287
    |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 440 CTTTCACCTGTAGAAACCATTAATGTGCCAATAGTTAGTTCGAATATAGCAGGTAA 381
    |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 288 AlaValGlyLeuProPheAspAlaMetLeuGlyAlaLeuLysAlaPheSerGlyLeuAla 307
    |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 380 CTGCTAAATATAGTAATGAAAGTATCCGCAATAGTCTGTATGACATTTTCAGCGTTCG 321
    |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 308 HisArgCysGlnTrpValArgGluArgGlnGlyValSerTyrAspSerLysAla 327
    |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 320 CACAGATTGGAGCAGGTTCGAGTTATTATATGATGTGAATACATCAATGACAGTAAGGCA 261
    |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 328 ThrAsnValGlyAlaAlaLeuAlaIleGluGlyLeuGlyAlaAspIleAspGlyLys 347
    |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 260 ACCAATGTAATGTGCTATATATGCACTGAA-----AGTGTGAATATCTCT 213
    |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 348 LeuValLeuLeuAlaGlyAspGlyLysGlyAlaAspPheHisAspLeuArgGluPro 367
    |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 212 GCAATATGATTTGTTGGGAATAGATAAGAAAGAAATGACTATACAGAAATTCAGGATCG 153
    |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 368 ValAlaArgPheCysArgAlaValValLeuLeuGlyArgAspAlaGlyLeuIleAlaGln 387
    |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 152 GTAAAGAAAGAAAGTTCGTGCCATCGTATGCTGGGATTGGTAATGAAAGATCATTCAG 93
    |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 388 AlaLeuGlyAsnAlaValProLeuVal---ArgValAlaThrLeuAspGluAlaVal 405
    |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 92 TTCCTCAGAAACAAAAGGATTTGATTTTGAACCTCTAGCATGGAAGAAATGTGTA 36
    |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

```

## RESULT 9

BZ550283/c 1194 bp DNA linear GSS 17-DEC-2002  
LOCUS pasci-60\_2584.xl pasci-60 Pseudomonas aeruginosa genomic clone  
DEFINITION pasci-60\_2584, genomic survey sequence.  
ACCESSION BZ550283

VERSION BZ550283.1 GI:27153864  
KEYWORDS GSS.  
SOURCE Pseudomonas aeruginosa

ORGANISM Pseudomonas aeruginosa  
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
Pseudomonadaceae; Pseudomonas.

## REFERENCE

1 (bases 1 to 1194)  
Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,  
Burns, J.L., Kaul, R. and Olsen, M.V.  
Whole-Genome-Sequence variation among multiple isolates of  
Pseudomonas aeruginosa library  
J. Bacteriol., (2002) In press  
Contact: Chris K. Raymond  
Genome Center  
University of Washington  
Box 352145, Seattle, WA 98105-2145, USA  
Tel: 2062216954  
Fax: 2066857244  
Email: craymond@u.washington.edu  
Class: shotgun.

## FEATURES

source  
1..1194  
Location/Qualifiers  
/organism="Pseudomonas aeruginosa"  
/mol\_type="genomic DNA"  
/strain="1-60"  
/db\_xref="taxon:287"  
/clone="pasci-60\_2584"  
/clone\_lib="pasci-60"  
/note="clinical isolate 1-60 whole genomic shotgun  
library."

BASE COUNT 233 a 349 c 298 g 314 t  
ORIGIN

Alignment Scores:  
Pred. No.: 1.04e-14 Length: 1194  
Score: 226.00 Matches: 77  
Percent Similarity: 54.72% Conservative: 10  
Best Local Similarity: 48.43% Mismatches: 51  
Query Match: 10.00% Indels: 23  
DB: 29 Gaps: 2

US-09-701-229-2 (1-448) x BZ550283 (1-1194)

```

QY 311 GlnTrpValArgGluArgGlnGlyValSerTyrTyrAspAspSer----- 325
    |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 884 AGGTGGAGACAGAAAAGAAATTTATTGATTAGACTAGATTCCAGCCAGCCGGGCCA 825
    |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 326 -----LysAlaThrAsnValGlyAlaAlaLeu----- 334
    |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 824 CCGGACGGGAACAAAAGCAAAACGAGGTAGTGCATGTGTCAACAAACGGTGTCA 765
    |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 335 -----AlaAlaIleGluGlyLeuGlyAlaAspIleAspGlyLysLeuValLeu 350
    |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 764 TAAACCTTTCAGCCGGATCAAGAGAGATATATACCTACTTTTCGCCAGCGATGTGTC 705
    |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 351 Leu-AlaGlyAspGlyLysGlyAlaAspPheHisAspLeuArgGluProValAlaAr 370
    |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 704 CTCGCCGGAGAGAGGACACGCCACCATTTTCATGTCTTTGCCGACCGCGTTC 645
    |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 370 gPheCysArgAlaValValLeuLeuGlyArgAspAlaGlyLeuIleAlaGlnAlaLeuGl 390
    |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 644 GTTTAA-CGAGCTGTGCTCTGTTTC-CGGGCCACCATCATTAATCCCGAGCGGTGG 587
    |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 390 YAsnAlaValProLeuValArgValAlaThrLeuAspGluAlaValArgGlnAlaAlaGl 410
    |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 586 CAACACGGCTCATTTGGTGGCGCTCGCAGCTTCGGACGAAGCAGTTCGGCGAGCGCCTA 527
    |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 410 uLeuAlaArgGluGlyAspAlaValLeuLeuSerProAlaCysAlaSerLeuAspMetPh 430
    |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 526 CTTGCCCTCAAGACAGATGCGGTGCTGTTGTAGCCGCCCTGCGGAGACCTGTACATGTT 467
    |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 430 eLysAsnPheGluGluArgGlyArgLeuPheAlaLysAlaValGluGluLeuAla 448
    |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

```



Db 466 CAAGAACTTCGAAGACCGCAACGCTGTCGCCAAACCGTAGAGGAGCTAGCG 412

RESULT 10  
BH795426/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

BH795426 572 bp DNA linear GSS 09-APR-2002  
BMBAC376G09P7 PSU Brugia malayi Genomic Bac Library 3 Brugia malayi  
genomic; genomic survey sequence.

BH795426  
BH795426.1 GI:20043758  
GSS.

Brugia malayi  
Brugia malayi  
Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;  
Onchocercidae; Brugia.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

1 (bases 1 to 572)  
Whitton,C., Daub,J., Ware,J., Quail,M., Hall,N., Barrell,B., Foster  
J., Guilianio,D., Slatko,B. and Blaxter,M.  
Genome survey sequences from the human parasitic nematode Brugia  
malayi  
Unpublished  
Contact: Blaxter ML  
Institute of Cell, Animal and Population Biology  
University of Edinburgh  
Ashworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9  
3JT, UK  
Tel: +44 131 650 6760  
Fax: +44 131 670 5450  
Email: mark.blaxter@ed.ac.uk  
Sequenced from the Brugia malayi BAC library constructed by Claire  
Whitton and Dr Mike Quail. The sequence was generated by The  
Pathogen Sequencing Unit, The Sanger Institute, Cambridge, UK in  
collaboration with Mark Blaxter, ICAPB, University of Edinburgh,  
Edinburgh, UK  
Seq primer: T7 (TAATACGACTCACTATAGG)  
Class: BAC ends.

FEATURES  
source  
Location/Qualifiers  
1..572  
/organism="Brugia malayi"  
/mol\_type="genomic DNA"  
/strain="TRS"  
/db\_xref="taxon:6279"  
/sex="Mixed (male and female)"  
/tissue\_type="whole parasite"  
/dev\_stage="microfilaria (L1)"  
/clone\_lib="Brugia malayi Genomic Bac Library 3"  
/note="Vector: pBACe3.6; Site 1: BamH I; Brugia malayi  
genomic DNA was partially cleaved with Sau3A I and size  
fractionated. 7,392 clones were generated with mean insert  
size ~48 kbp. The library was constructed by Claire  
Whitton, Blaxter Nematode Genetics Lab, University of  
Edinburgh, UK, and Dr Mike Quail, The Pathogen Sequencing  
Unit, The Sanger Centre, Cambridge, UK."

BASE COUNT 154 a 120 c 86 g 212 t

ORIGIN

Alignment Scores:  
Pred. No.: 8,84e-14 Length: 572  
Score: 213.50 Matches: 56  
Percent Similarity: 61.1% Conservative: 21  
Best Local Similarity: 44.4% Mismatches: 31  
Query Match: 9.45% Indels: 18  
DB: 28 Gaps: 6

US-09-701-229-2 (1-448) x BH795426 (1-572)

QY 52 TyrProGlnValGluValArgCysGly-GluLeuAspAlaGluPheLeuCysSerAlaAr 71  
||||| : : : : :  
Db 504 TATCCATCAAAAATAGATTGGCAGAGATAAAGCA----- 465

QY 71 gGluLeuTyValSerProGlyLeuSerLeuArgThrProAla-----LeuValGl 88  
||| : : : : :  
Db 464 ----TTGATTTTAAGCCCTGGAGTACCAATTTTCATATCCGAGCCACATTTGGATAGTAA 409

QY 88 nAlaAlaAlaLys---GlyValArgIleSerGlyAspIleAspLeuPheAlaArgGluAl 107  
: ||| : : : : :  
Db 408 ACTTGCAGAAGATTTTCTGACTGTAAAAATAAAATCGGATATTGAACTATTCTTA---GAAGC 352  
||| : : : : :  
QY 107 aLysAla-----ProIleValAlaIleThrGlySerAspAlaLysSerThrValTh 124  
||||| : : : : :  
Db 351 TAAAGCTAAGACACGAGAAATTTGTAGCGCTCAGAGAACGAATGGCAAAATCAACCACTAC 292  
||||| : : : : :  
QY 124 rThrLeuValGlyGluMetAlaValAlaAlaAspLysArgValAlaValAlaValGlyGlyAsnLe 144  
: : : : :  
Db 291 GTCAATTAATAGTTCATATTAATCTCGGGGAAAAAAGTAGCTATTGGTGAATTT 232  
: : : : :  
QY 144 uGlyThrProAlaLeuAspLeuAlaAspIleGluLeuTyValLeuGluLeuSe 164  
||| : : : : :  
Db 231 AGGTGTTCTATTTTTGGAT---CTAGAAAAAGATCGGAAATTTATGTAAATTCAAATTC 175  
||| : : : : :  
QY 164 rSerPheGlnLeuGlu 169  
||||| : : : : :  
Db 174 CTCCTTTCAATTAGAA 159  
||||| : : : : :

RESULT 11  
CNS070MR/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

CNS070MR 965 bp DNA linear GSS 06-JUL-2001  
T3 end of clone AZ0AA010F07 of library AZ0AA from strain CBS 712 of  
Kluyveromyces marxianus, genomic survey sequence.

AL423817  
AL423817.1 GI:12207011  
GSS.

Kluyveromyces marxianus  
Kluyveromyces marxianus  
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
Saccharomycetales; Saccharomycetaceae; Kluyveromyces.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

1 (bases 1 to 965)  
Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,  
Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,  
de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Liorente,B.,  
Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,  
Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,  
Wincker,P. and Weissenbach,J.  
Genomic exploration of the hemiascomycetous yeasts: 1. A set of  
yeast species for molecular evolution studies  
FEBS Lett. 487 (1), 3-12 (2000)  
20584711  
11152876

2 (bases 1 to 965)  
Llorente,B., Malpertuy,A., Blandin,G., Artiguenave,F., Wincker,P.  
and Dujon,B.  
Genomic exploration of the hemiascomycetous yeasts: 12.  
Kluyveromyces marxianus var. marxianus  
FEBS Lett. 487 (1), 71-75 (2000)  
20584722  
11152887

3 (bases 1 to 965)  
Genoscope.  
Direct Submission  
Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,  
2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :  
seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)  
This GSS is part of a random genomic sequencing program of thirteen  
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces  
exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,  
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces  
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia  
angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,  
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to  
5 kb were prepared and both extremities were sequenced. See  
keywords for description of this sequence and for the sequence of  
the other extremity of this insert.

FEATURES  
source  
Location/Qualifiers  
1..965  
/organism="Kluyveromyces marxianus"  
/mol\_type="genomic DNA"  
/strain="CBS 712"



[illegible]

Db 144 CACCTTGGCGGCGATGTCGAGACCGTCCATAAGCTGGTGGCATCACCTGGGTGCGAC 203  
 QY 324 AspSerLysAlaThrAsnValGlyAlaAlaLeuAlaAlaLeuAlaAlaLeuGlyAlaAlaAsp 343  
 Db 204 GACTCCAGCGCCACCAACCTCAGCGCGGAACTCTCGATCGGCGCTTC----- 254  
 QY 344 IleAspGlyLysLeuValLeuLeuAlaGlyGlyAspGlyLysGlyAlaAlaAspPheHisAsp 363  
 Db 255 -----GAGCACATCGTGTGGATMGCCGAGGAGCAG-GCCAAGGGCCACCACTTTGACGAC 307  
 QY 364 Leu-ArgGluProValAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 383  
 Db 308 CTGGTCACCAACCGCCGAGAGCTGCGCGNAGCATCGTG---CTGGGAACCGACCGTG 364  
 QY 383 LysLeuAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 403  
 Db 365 AGCTCKTCGCCAATHTCTTTCCGAGCAGCGCCGCAATTCCTCCGTCGTCATCGGAG 424  
 QY 403 lu-----AlaValArgGlnAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 415  
 Db 425 ACATTTNGCGAGAGTGCATGCGCAAGGCGAGTNCGCGAGGCTCCGAGATGGCCCGCG 484  
 QY 415 LysAlaValLeuLeuSerProAlaCysAlaSerLeuAspMetPheLysAsnPheGluG 435  
 Db 485 GCGACACCGCTCTGATGCTCGCGGTCGCGGACATC-GACATTTGGCGTGGATATGCTG 543  
 QY 435 luArgGlyArgLeuPheAlaLysAlaValGluGluLeu 447  
 Db 544 CCCGCGGTGACGATTCGTCACAGCGCGCGACGACAAATT 581

## RESULT 13

BH383040/c

LOCUS

DEFINITION BH383040 616 bp DNA linear GSS 10-DEC-2001  
 AG-ND-127115.TF ND-TAM Anopheles gambiae genomic clone AG-ND-127115  
 , genomic survey sequence.

ACCESSION

BH383040

VERSION

BH383040.1 GI:17329182

KEYWORDS

GSS.

SOURCE

ORGANISM

Anopheles gambiae (African malaria mosquito)

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;

Anopheles.

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

Other GSSs: AG-ND-127115, TF

Contact: Brendan J Loftus

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0208

Fax: 301 838 3543

Email: bjloftus@tigr.org

This clone is from an A. gambiae BAC library (ND-TAM) provided by

F.H. Collins and sequenced by The Institute for Genomic Research

(TIGR). The BAC library was generated from A. gambiae PEST strain

DNA. All DNA was extracted from newly hatched first instar larvae

to minimize the inclusion of DNA from microorganisms that inhabit

the gut. The DNA is derived from mixed sexes of larvae. The BAC

library was constructed at Texas A&amp;M University BAC Center

University, College Station, Texas 77843-2123, USA using a HindIII

partial digest.

Seq primer: M13 Rev

Class: BAC ends.

## FEATURES

source

Location/Qualifiers

1. .616

/organism="Anopheles gambiae"

/mol\_type="genomic DNA"

/strain="PEST"

/db\_xref="taxon:7165"

/clone\_lib="ND-127115"

/note="Vector: pECBAC1; Site\_1: HindIII"

BASE COUNT 182 a 134 c 87 g 213 t

ORIGIN

Alignment Scores:

Pred. No.: 1.3e-10

Score: 186.00

Percent Similarity: 50.34%

Best Local Similarity: 31.23%

Query Match: 8.23%

Indels: 4

Gaps: 1

US-09-701-229-2 (1-448) x BH383040 (1-616)

QY 250 AspGlyGlnLysTrpLeuAlaPheGlnPheAspLysLeuProValGlyGluLeuLys 269

Db 471 GACAATGAATAGTAGTAAACTTCAGGACGACATTCAGAATGAAAAATTTCTGATCTTCA 412

QY 270 IleArgGlyAlaHisAsnTy-SerAsnAlaLeuAlaAlaLeuAlaLeuGlyHisAlaVal 289

Db 411 CTGTGAGNAACCAATAATGTTGCCAATAGTTAGCTGCAAGTATAGCAGGTAACCTGCTA 352

QY 290 GlyLeuProPheAspAlaMetLeuGlyAlaLeuLysAlaPheSerGlyLeuAlaHisArg 309

Db 351 AATATTAGTAATGAAGATATCGCAATAGTCTGATGACATTTTCAGGCTGTTCCGCACAGA 292

QY 310 CysGlnTrpValArgGluArgGlnGlyValSerTy-TyrAspAspSerLysAlaThrAsn 329

Db 291 TTGGAGCAGGTTCAGTATTAAATGATGTAATCAATCAATGACAGTAAGCAACCAAT 232

QY 330 ValGlyAlaAlaLeuAlaAlaLeuGlyLeuGlyAlaAlaAspLysLeuVal 349

Db 231 GTAAATGCTGCATATTATGCTACTGCA-----AGTGAATAATATCTGCAATA 184

QY 350 LeuLeuAlaGlyGlyAspGlyLysGlyAlaAlaPheHisAspLeuArgGluProValAla 369

Db 183 TGGATTGTTGGTGGAAATAGATAAGGAAATGACTATACAGAAATTCAGGATCTGCTGAAG 124

QY 370 ArgPheCysArgAlaValValLeuLeuGlyArgAspAlaGlyLeuIleAlaGlnAlaLeu 389

Db 123 AAGAAAGTTCGTGCCATCGTATGCTGGGATGGATTAATGAAAGATCATTTCAGTCTTC 64

QY 390 GlyAsnAlaValProLeuVal 396

Db 63 AGAAACAAAAGGATTTGATT 43

RESULT 14

BH369957/c

LOCUS

DEFINITION

AG-ND-138K1.TF ND-TAM Anopheles gambiae genomic clone AG-ND-138K1,

genomic survey sequence.

ACCESSION

BH369957

VERSION

BH369957.1 GI:17316060

KEYWORDS

GSS.

SOURCE

ORGANISM

Anopheles gambiae (African malaria mosquito)

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;

Anopheles.

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

Other GSSs: AG-ND-127115, TF

Contact: Brendan J Loftus

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0208

Fax: 301 838 3543

Email: bjloftus@tigr.org

This clone is from an A. gambiae BAC library (ND-TAM) provided by

F.H. Collins and sequenced by The Institute for Genomic Research

(TIGR). The BAC library was generated from A. gambiae PEST strain

DNA. All DNA was extracted from newly hatched first instar larvae

to minimize the inclusion of DNA from microorganisms that inhabit

the gut. The DNA is derived from mixed sexes of larvae. The BAC

library was constructed at Texas A&amp;M University BAC Center

University, College Station, Texas 77843-2123, USA using a HindIII

partial digest.

Seq primer: M13 Rev

Class: BAC ends.

```

JOURNAL      malaria mosquito Anopheles gambiae
MEDLINE      Mol. Genet. Genomics 268 (6), 720-728 (2003)
PUBMED      22542063
COMMENT      12655398
              Contact: Brendan J Loftus
              Department of Eukaryotic Genomics
              The Institute for Genomic Research
              9712 Medical Center Dr., Rockville, MD 20850, USA
              Tel: 301 838 0208
              Fax: 301 838 3543
              Email: bjloftus@tigr.org
              This clone is from an A. gambiae BAC library (ND-TAM) provided by
              F.H. Collins and sequenced by The Institute for Genomic Research
              (TIGR). The BAC library was generated from A. gambiae PEST strain
              DNA. All DNA was extracted from newly hatched first instar larvae
              to minimize the inclusion of DNA from microorganisms that inhabit
              the gut. The DNA is derived from mixed sexes of larvae. The BAC
              library was constructed at Texas A&M University BAC Center
              University, College Station, Texas 77843-2123, USA using a HindIII
              partial digest.
              Seq primer: M13 For
              Class: BAC ends.

FEATURES             source
    Location/Qualifiers
        1..495
        /organism="Anopheles gambiae"
        /mol_type="genomic DNA"
        /strain="PEST"
        /db_xref="taxon:7165"
        /clone="AG-ND-138K1"
        /clone_lib="ND-TAM"
        /note="vector: pECBAC1; Site_1: HindIII"

BASE COUNT      143 a 110 c 68 g 174 t
ORIGIN

Alignment Scores:
Pred. No.:      3.15e-09      Length:      495
Score:          172.50      Matches:      43
Percent Similarity: 52.24%      Conservative: 27
Best Local Similarity: 32.09%      Mismatches: 59
Query Match:    7.63%      Indels:      5
DB:             28      Gaps:      2

US-09-701-229-2 (1-448) x BH369957 (1-495)

Qy 274 HisAsnTyrSerAsnAlaLeuAlaLeuAlaLeuGlyHisAlaValGlyLeuProPhe 293
||||| :||||:||||| :||| :||| :|||
Db 392 CATATGTGCCATAGTTTGTAGCTGCAAGTATACGAGTAACTGCTAAATATTAGTAAT 333
||||| :||||:||||| :||| :||| :|||

Qy 294 AspAlaMetLeuGlyAlaLeuLysAlaPheSerGlyLeuAlaHisArgCysGlnTrpVal 313
||||| :||||:||||| :||| :||| :|||
Db 332 GAAAGTATCCGCAATAGCTGATGACATTTTCAGGCTGTCGCGACAGATTGGACAGGTT 273
||||| :||||:||||| :||| :||| :|||

Qy 314 ArgGluArgGlnGlyValSerTyrTyrAspSerLysAlaThrAsnValGlyAlaAla 333
||||| :||||:||||| :||| :||| :|||
Db 272 GCAGTTATTATGATGGAATCATCATGACAGTAAAGCAACCAATGTAATATGCTGCA 213
||||| :||||:||||| :||| :||| :|||

Qy 334 LeuAlaAlaLeuGlyLeuGlyAlaAspLysLeuValLeuLeuAlaGly 353
||||| :||||:||||| :||| :||| :|||
Db 212 TATTATGCATGGAA-----AGTGTGAATATCTCGCAATATGATGTTTGGT 165
||||| :||||:||||| :||| :||| :|||

Qy 354 GlyAspGlyLysGlyAlaAspPheHisAspLeuArgGluProValAlaArgPheCysArg 373
||||| :||||:||||| :||| :||| :|||
Db 164 GGAATAGATAAGGAAATGACTATACAGAAATTCAGGATCTGGTAAAGAAAGTTCGT 105
||||| :||||:||||| :||| :||| :|||

Qy 374 AlaValValLeuLeuGlyArgAspAlaGlyLeuIleAlaGlnAlaLeuGlyAsnAlaVal 393
||||| :||||:||||| :||| :||| :|||
Db 104 GCCATCGTATGCTCGGGATGGAATAATGAAAGATCATTCAGTCTTCACAGAAACAAAAG 45
||||| :||||:||||| :||| :||| :|||

Qy 394 ProLeuVal---ArgValAlaThrLeuAspGluAlaValArg 406
||||| :||||:||||| :||| :||| :|||
Db 44 GATTGTATTTTGAACCTCTAGCATGGAAGATGCTGTAAG 3
||||| :||||:||||| :||| :||| :|||

RESULT 15

```

```

BH382726/c
LOCUS      BH382726      378 bp      DNA      linear      GSS 10-DEC-2001
DEFINITION AG-ND-127B23_TR ND-TAM Anopheles gambiae genomic clone AG-ND-127B23
            , genomic survey sequence.
ACCESSION  BH382726
VERSION    BH382726.1 GI:17328868
KEYWORDS   GSS.
SOURCE     Anopheles gambiae (African malaria mosquito)
ORGANISM   Anopheles gambiae
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea,
            Anopheles.
REFERENCE  1 (bases 1 to 378)
AUTHORS   Hong,X.S., Hogan,J.R., Wang,X., Sarkar,A., Sim,C., Loftus,B.J., Ren
            ,C., Huff,E.R., Carlile,J.L., Black,K., Zhang,H.-B., Gardner,M.J.
            and Collins,F.H.
TITLE     Construction of a BAC library and generation of BAC end
            sequence-tagged connectors for genome sequencing of the African
            malaria mosquito Anopheles gambiae
JOURNAL    Mol. Genet. Genomics 268 (6), 720-728 (2003)
MEDLINE    22542063
PUBMED     12655398
COMMENT    Contact: Brendan J Loftus
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0208
            Fax: 301 838 3543
            Email: bjloftus@tigr.org
            This clone is from an A. gambiae BAC library (ND-TAM) provided by
            F.H. Collins and sequenced by The Institute for Genomic Research
            (TIGR). The BAC library was generated from A. gambiae PEST strain
            DNA. All DNA was extracted from newly hatched first instar larvae
            to minimize the inclusion of DNA from microorganisms that inhabit
            the gut. The DNA is derived from mixed sexes of larvae. The BAC
            library was constructed at Texas A&M University BAC Center
            University, College Station, Texas 77843-2123, USA using a HindIII
            partial digest.
            Seq primer: M13 Rev
            Class: BAC ends.

FEATURES             source
    Location/Qualifiers
        1..378
        /organism="Anopheles gambiae"
        /mol_type="genomic DNA"
        /strain="PEST"
        /db_xref="taxon:7165"
        /clone="AG-ND-127B23"
        /clone_lib="ND-TAM"
        /note="vector: pECBAC1; Site_1: HindIII"

BASE COUNT      108 a 85 c 52 g 133 t
ORIGIN

Alignment Scores:
Pred. No.:      2.38e-09      Length:      378
Score:          172.00      Matches:      42
Percent Similarity: 52.00%      Conservative: 23
Best Local Similarity: 33.60%      Mismatches: 56
Query Match:    7.61%      Indels:      4
DB:             28      Gaps:      1

US-09-701-229-2 (1-448) x BH382726 (1-378)

Qy 267 GluLeuLysIleArgGlyAlaHisAsnTyrSerAsnAlaLeuAlaLeuAlaLeuGly 286
||||| :||||:||||| :||| :||| :|||
Db 375 GATCTTCACACGTAGGAAACCAATAATGTTCCCAATAGTTTACGCGAAGTATACGAGT 316
||||| :||||:||||| :||| :||| :|||

Qy 287 HisAlaValGlyLeuProPheAspAlaMetLeuGlyAlaLeuLysAlaPheSerGlyLeu 306
||||| :||||:||||| :||| :||| :|||
Db 315 AAACGTCTAAATATTAGTAAAGTATCCGAATAGCTGATGACATTTTCAGGCTGTT 256
||||| :||||:||||| :||| :||| :|||

Qy 307 AlaHisArgCysGlnTrpValArgGluArgGlnGlyValSerTyrTyrAspSerLys 326
||||| :||||:||||| :||| :||| :|||
Db 255 CCGCACAGATTGGACAGGTTGCAGTTATTATGATGTGAAATACATCATCATGACAGTAAG 196
||||| :||||:||||| :||| :||| :|||

```

QY	327	AlaThrAsnValGlyAlaAlaLeuAlaAlaIleGluGlyLeuGlyAlaAspIleAspGly	346
Db	195	GCAACCAATGTAATGTCATATTTATGCACTGAA-----AGTGTGAATAT	148
QY	347	LysLeuValLeuLeuAlaGlyGlyAspGlyLysGlyAlaAspPheHisAspLeuArgGlu	366
Db	147	CCTGCAATATGGATTGTTGGTGAATAGATAAAGGAATGACTATACAGAAATTGAGGAT	88
QY	367	ProValAlaArgPheCysArgAlaValLeuLeuGlyArgAspAlaGlyLeuIleAla	386
Db	87	CTGTAAGNAGAAAGTTCTGCCATCGTATGTCGGATTGGATAATGAAAAGATCATT	28
QY	387	GlnAlaLeuGlyAsn	391
Db	27	CAGTTCTTCAGAAAC	13

Search completed: August 14, 2003, 09:10:32  
 Job time : 2302 secs

TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: mRNA  
 HYPOTHEICAL: NO  
 ANTI-SENSE: NO  
 US-08-843-309-3

Alignment Scores:  
 Pred. No.: 1,86e-55 Length: 1350  
 Score: 586.50 Matches: 162  
 Percent Similarity: 52.92% Conservativity: 83  
 Best Local Similarity: 34.99% Mismatches: 177  
 Query Match: 25.95% Indels: 41  
 DB: 2 Gaps: 14

US-09-701-229-2 (1-448) x US-08-843-309-3 (1-1350)

```

Qy 7 AspHisPheArg-----IleValValGlyLeuGlyLysSerGlyMetSerLeu 22
Db 13 GAUCAAUUUAAGAAAGAAAGUUCUUUAGGUUGGCCAAGUCUGGUGAAUCUGCA 72

Qy 23 ValArgTyrLeuAlaArgArgGlyLeuProPheAlaValValAspThrArg----- 39
Db 73 GCUCGUUUUGGACAAAGCUGGUGCCAUUGUGACAGAAUAGGAAACCUUUCGAG 132

Qy 40 GluAsnProGluLeuAlaThrLeuArgAlaGlnTyrProGlnValGluValArgCys 59
Db 133 GACAUAUCCAGCUGCCCAAGUUUCUGGAA-----GNAAGGAUCCAGGUAUACA 183

Qy 60 GlyGlu-----LeuAspAlaGluPheLeuCysSerAlaArgGluLeuTyr 74
Db 184 GGUGCCCAUCCUUGGAACUCUUGGAAGAUUUGCCUUAUUGGUGAAA----- 234

Qy 75 ValSerProGlyLeuSerLeuArgThrProAlaLeuValGlnAlaAlaLysGlyVal 94
Db 235 ---AAUCCAGUAUCCUCCUACAAUCCCAUCCCAUCCCAUCCCAUCCCAUCCCAU 291

Qy 95 ArgIleSerGlyAspIleAspLeuPheAlaArgGluAlaLysAlaProIleValAlaIle 114
Db 292 CCAGUCUUGACUGAGGUGGAUUGGUUUUUGAUUUCAGAGCAGCAGCAGCAGCAGCAG 351

Qy 115 ThrGlySerAsnAlaLysSerThrValThrLeuValGlyGluMetAlaValAlaAla 134
Db 352 ACAGGAUCCAGGUAAGACACACACACACACACACACACACACACACACACACACAC 411

Qy 135 AspLysArgValAlaValGlyGlyAsnLeuGlyThrProAlaLeuAspLeu-----Leu 152
Db 412 GGGCAACAUGGUCUUUUAUCCAGGGAUUAUCCGCUUUAUCCGCUUUAUCCGCUUUA 471

Qy 153 AlaAspAspIleGluLeuTyrValLeuGluLeuSerSerPheGlnLeuGluThrCysAsp 172
Db 472 GCAUCAGAAAGGACACGCUUUGUAUCCUUCUUCUUCUUCUUCUUCUUCUUCUUCU 531

Qy 173 ArgLeuAsnAlaGluValAlaThrValLeuAsnValSerGluAspHisMetAspArgTyr 192
Db 532 GAUUAUCCAGCAGAUUGGUGGUUAUUAUCCCAUCCCAUCCCAUCCCAUCCCAUCCCA 591

Qy 193 AspGlyMetAlaAspTyrHisLeuAlaLysHisArgIle-----PheArgGlyAla 209
Db 592 GGGUUAUUCUGAAUUAUUGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 651

Qy 210 ArgGlnValValValAsnArgAlaAspAlaLeuThrArgProLeuIleAlaAspThr--- 228
Db 652 GAUUAUCCUUGUUGAAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 711

Qy 229 -----ValProCysTrpSerPheGlyLeuAsnLysProAspPheLysAlaPheGly 245
Db 712 GCCACUGUUGUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 753

Qy 246 LeuIleGluGluAspGlyGlnLysTrpLeuAlaPheGlnPheAspLysLeuLeuProVal 265
Db 754 GCUUAUCCUGGAAGAUGGUGCAA-----CUCUUAUCCUGGUGGAAGUAGUAGUAG 807

```

```

Qy 266 GlyGluLeuLysIleArgGlyAlaHisAsnTyrSerAsnAlaLeuAlaLeuAlaLeu 285
Db 808 AAUGAAAUCCGUGUUGUCCAGGAGCCACAAGUGGAAAUAGCCUUGGCGAUUUGCUGUA 867

Qy 286 GlyHisAlaValGlyLeuProPheAspAlaMetLeuGlyAlaLeuLysAlaPheSerGly 305
Db 868 GCCAAGCUGUGAUGGUGGACAAUCAAACCAAGGAAACUUAUCCGCUUUGGUGGUG 927

Qy 306 LeuAlaHisArgCysGlnTrpValArgGluArgGlnGlyValSerTyrTyrAspSer 325
Db 928 GUCAAAACACCGUCCAGUUUGUGGAGCAGCAUCAAAGGUGUUAUUAUUAUUAUUA 987

Qy 326 LysAlaThrAsnValGlyAlaAlaLeuAlaAlaIleGluGlyLeuGlyAlaAspIleAsp 345
Db 988 AAUCAACUAUAUCCUUGGUCUACUAAACCAACCUUAUCCAGGAUUGAC-----AAC 1038

Qy 346 GlyLysLeuValLeuLeuAlaGlyGlyAspGlyLysGlyAlaAspPheHisAspLeuArg 365
Db 1039 AGCAAGGCGUCUUGAUGGAGGUGUUGGAGGUGUUGGAGGUGUUAUUAUUAUUA 1095

Qy 366 GluProValAlaArgPheCysArgAlaValValLeuLeuGlyArgAspAlaGlyLeuIle 385
Db 1096 GUGCCAGACAUUCCUGGACUACAGAAUGGUCUCCUGGUGUUAUUAUUAUUAUUA 1155

Qy 386 AlaGlnAlaLeuGlyAsnAla---ValProLeuValArgValAlaThrLeuAspGluAla 404
Db 1156 AAACGGGCGACAGACAGGCGUGGUGUCCUUAUUGGAGGCGCAGAGAUUUGCAGACG 1215

Qy 405 ValArgGlnAlaAlaGluLeuAlaArgGluGlyAspAlaValLeuLeuSerProAlaCys 424
Db 1216 ACCGCGAAGGCGUAGAGCUGGUCGACUACAGGAGAGUUGUUCUUAUUAUUAUUA 1275

Qy 425 AlaSerLeuAspMetPheLysAsnPheGluGluArgGlyArgLeuPheAlaLysAlaVal 444
Db 1276 GCCAGCUGGGAUUGAUUGUUAUCCUUAUUAUUAUUAUUAUUAUUAUUAUUAU 1335

Qy 445 GluGluLeu 447
Db 1336 GCGGAGUUA 1344

```

# RESULT 15

```

US-08-843-309-1
; Sequence 1, Application US/08843309
; Patent No. 5834270
; GENERAL INFORMATION:
; APPLICANT: Skatrud, Paul
; APPLICANT: Peery, Robert
; APPLICANT: Hoskins, JoAnn
; APPLICANT: Wu, Chyun-Yeh Earnest
; TITLE OF INVENTION: Biosynthetic Gene Mur D of Streptococcus pneumoniae
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: US
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/08/843,309
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Webster, Thomas D
; REGISTRATION NUMBER: 39,872
; REFERENCE/DOCKET NUMBER: X-9900
; TELECOMMUNICATION INFORMATION:

```

TELEPHONE: 317-276-3334  
 TELEFAX: 317-276-3861  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1353 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 HYPOTHETICAL: NO  
 ANTI-SENSE: NO  
 FEATURE:   
 NAME/KEY: CDS  
 LOCATION: 1..1353  
 US-08-843-309-1

*Sequence!*  
*of*  
*US Pat. 5,834,270*

Alignment Scores:  
 Pred. No.: 1 87e-55 Length: 1353  
 Score: 586.50 Matches: 162  
 Percent Similarity: 52.92% Conservative: 83  
 Best Local Similarity: 34.99% Mismatches: 177  
 Query Match: 25.95% Indels: 41  
 DB: 2 Gaps: 14

US-09-701-229-2 (1-448) x US-08-843-309-1 (1-1353)

QY 7 AspHisPheArg-----lLeValValGlyLeuGlyLysSerGlyMetSerLeu 22  
 DB 13 GATCAATTAATAAAGAAAGTCTCTTTAGTTTGGCCCAAGTCTGGTGAATCTGCA 72  
 QY 23 ValArgTyrLeuAlaArgGlyLeuProPheAlaValValAspThrArg----- 39  
 DB 73 GCTCGTTTGTGGACAAGTAGTGCCATTGTGACAGTAATGATGGGAAACCTTTCGAG 132  
 QY 40 GluAsnProGluLeuAlaThrLeuArgAlaGlnTyrProGlnValGluValArgCys 59  
 DB 133 GACAATCCAGCTGCCCAAGTTTGTGGAA-----GAAGGGATCAAGTCTATTACA 183  
 QY 60 GlyGlu-----LeuAspAlaGluPheLeuCysSerAlaArgGluLeuTyr 74  
 DB 184 GTGGGCCATCCCTTTGGAACCTCTTGGATGAAGAGTTTGCCTTATGGTGAAA----- 234  
 QY 75 ValSerProGluLeuSerLeuArgThrProAlaLeuValGlnAlaAlaLysGlyVal 94  
 DB 235 ---ATCCAGGTATCCCTACAAATCCATGATGAAAGGCTTGGCCCAAGAGAAT 291  
 QY 95 ArgIleSerGlyAspIleAspLeuPheAlaArgGluAlaLysAlaProIleValAlaIle 114  
 DB 292 CCAGTCTTCTGCTGAGTGGAATTTGCTTATTTGATTTTCAAGAGCAGCGATTATGATC 351  
 QY 115 ThrGlySerAsnAlaLysSerThrValThrLeuValGlyGluMetAlaAlaAla 134  
 DB 352 ACAGGATCGAAGCGTAAGAACCAACCAAGACTATGATTTGGGGAAGTTTGGTCTGCT 411  
 QY 135 AspLysArgValAlaValGlyAsnLeuGlyThrProAlaLeuAspLeu-----Leu 152  
 DB 412 GGGCAACATGGCTTTTATCAGGGAATATCGGCTATCTGCCAGTCAGGTCTCTCAATA 471  
 QY 153 AlaAspAspIleGluLeuTyrValLeuGluLeuSerSerPheGlnLeuGluThrCysAsp 172  
 DB 472 GCATCAGATAAGGACAGCGTCTTTATGGAACCTTCTTCTTCCACTCATGGGTGTTCAA 531  
 QY 173 ArgLeuAsnAlaGluValAlaThrValLeuAsnValSerGluAspHisMetAspArgTyr 192  
 DB 532 GAATTCATCCAGATGTCGGTTATTACCAACCTCATGCAACCTCATATCGACTACCAT 591  
 QY 193 AspGlyMetAlaAspTyrHisLeuAlaLysHisArgIle-----PheArgGlyAla 209  
 DB 592 GGGTCATTTCTGAATATGTAGCCCAAGTGGGAATATCCAGAACCAAGATGACACAGCT 651  
 QY 210 ArgGlnValValAlaValAsnArgAlaAspAlaLeuThrArgProLeuIleAlaAspThr--- 228  
 DB 652 GATTTCTTGTCTTGAACCTTTAATCAAGACTTGGCAAAAGACTTGCATCTCCAGACAGAA 711

QY 229 -----ValProCysTrpSerPheGlyLeuAsnLysProAspPheLysAlaPheGly 245  
 DB 712 GCCACTGTTGTACCACTTTTCAACA-----CTTGAAGAAGTTGAT-----GGA 753  
 QY 246 LeuIleGluGluAspGlyGlnLysTrpLeuAlaPheGlnPheAspLysLeuLeuProVal 265  
 DB 754 GCTTATCTGGAAGATGGTCAA-----CTCTACTCCGTGGTGAAGTAGTATGTCAGCG 807  
 QY 266 GlyGluLeuLysIleArgGlyAlaHisAsnTyrSerAsnAlaLeuAlaLeuAlaLeu 285  
 DB 808 AATGAATCGGTGTTCCAGGATGACCAATGTGGAAATGCCCTTCGCTACTATTGCTGTA 867  
 QY 286 GlyHisAlaValGlyLeuProPheAspAlaMetLeuGlyAlaLeuLysAlaPheSerGly 305  
 DB 868 GCCAAGCTTCGTGATGTGGACAATCAACCAATCAAGGAACTCTTTCACCCCTCGGTGT 977  
 QY 306 LeuAlaHisArgCysGlnTrpValArgGluArgGlnGlyValSerTyrTyrAspAspSer 325  
 DB 928 GTCAACACCCGCTCTCCAGTTTCTGGATGATCAACAGGGTGTAAATTTCTATAACGACAGT 987  
 QY 326 LysAlaThrAsnValGlyAlaAlaLeuAlaIleGluGlyLeuGlyAlaAspIleAsp 345  
 DB 988 AATCAACTAATATCTTGGCTACTCAAAAGGCTTATCAGGATTTCAC-----AAC 1038  
 QY 346 GlyLysLeuValLeuLeuAlaGlyAspGlyLysGlyAlaAspPheHisAspLeuArg 365  
 DB 1039 AGCAAGGTGCTCTTGTATTCAGGTGTTGGACCGTGGCAATGAGTTTACGCAATG--- 1095  
 QY 366 GluProValAlaArgPheCysArgAlaValValLeuLeuGlyArgAspAlaGlyLeuIle 385  
 DB 1096 GTGCCAGCATTTACTGGACTCAAGAGATGGTCTCTCTGGTCAATCTCGCAACGCTGTC 1155  
 QY 386 AlaGlnAlaLeuGlyAsnAla---ValProLeuValArgValAlaThrLeuAspGluAla 404  
 DB 1156 AAACGGCAGCAGACAGAGCTGCTGCTTATGTTGGAGCGCAGACATATTTCAGATGCG 1215  
 QY 405 ValArgGlnAlaAlaGluLeuAlaArgGluGlyAspAlaValLeuLeuSerProAlaCys 424  
 DB 1216 ACCCGCAAGGCCCTATGAGCTTCCGACTCAAGGAGATGTGTTCTTCTTCTGCTGCAAT 1275  
 QY 425 AlaSerLeuAspMetPheLysAsnPheGlyGluArgGlyArgLeuPheAlaLysAlaVal 444  
 DB 1276 GCCAGCTGGATATGTATGCTAACTTTGAAGTACGTGGCGACCTCTTTATCGACACAGTA 1335  
 QY 445 GluGluLeu 447  
 DB 1336 GCGGAGTTA 1344

Search completed: August 14, 2003, 09:37:52  
 Job time : 1636 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2003, 04:12:30 ; Search time 2898 seconds

(without alignments)  
12160.625 Million cell updates/sec

Title: US-09-701-229-1

Perfect score: 1450

Sequence: 1 cgtgctgacggcctcgcca.....tggtgagcggcgacggcgcac 1450

Scoring table: OLIGO\_NUC

Gapop 60.0 , Gapext 60.0

Searched: 22781392 seqs, 12152238056 residues

Word size : 0

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:\*

1: em\_estba:\*\*  
2: em\_esthum:\*\*  
3: em\_estin:\*\*  
4: em\_estmu:\*\*  
5: em\_estov:\*\*  
6: em\_estpl:\*\*  
7: em\_estro:\*\*  
8: em\_htc:\*\*  
9: gb\_est1:\*\*  
10: gb\_est2:\*\*  
11: gb\_htc:\*\*  
12: gb\_est3:\*\*  
13: gb\_est4:\*\*  
14: gb\_est5:\*\*  
15: em\_estfun:\*\*  
16: em\_estom:\*\*  
17: em\_gss\_hum:\*\*  
18: em\_gss\_inv:\*\*  
19: em\_gss\_pin:\*\*  
20: em\_gss\_vrt:\*\*  
21: em\_gss\_fun:\*\*  
22: em\_gss\_mam:\*\*  
23: em\_gss\_mus:\*\*  
24: em\_gss\_pro:\*\*  
25: em\_gss\_rod:\*\*  
26: em\_gss\_phg:\*\*  
27: em\_gss\_vrl:\*\*  
28: gb\_gss1:\*\*  
29: gb\_gss2:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
c 1	600	41.4	1445	29	BZ568959
c 2	510	35.2	892	29	BZ579116
c 3	125	8.6	1171	29	BZ559806
c 4	65	4.5	1194	29	BZ550283
					BZ550283 pacs1-60_

c 5	1754	10	BE966412
c 6	537	9	AL825430
c 7	687	10	BE586183
c 8	265	14	CB061196
c 9	369	9	AA969049
c 10	501	12	BI306480
c 11	792	10	BE041029
c 12	286	9	AW771204
c 13	289	14	CB042199
c 14	301	10	BF657011
c 15	318	9	AW139275
c 16	323	9	AW136188
c 17	338	13	BQ552547
c 18	376	10	BF109460
c 19	381	9	AI824553
c 20	397	9	AA013254
c 21	409	28	BH799135
c 22	440	10	BE633796
c 23	441	13	BY080930
c 24	472	9	AI291792
c 25	481	28	BH376496
c 26	489	13	BU040543
c 27	511	12	BQ286376
c 28	561	12	BI792784
c 29	571	12	BI792863
c 30	584	13	BU043073
c 31	589	12	BI792852
c 32	596	13	BQ269723
c 33	621	13	BU040487
c 34	633	9	AI819231
c 35	634	29	BX126826
c 36	665	9	AW191059
c 37	681	29	BZ781755
c 38	738	29	BX231314
c 39	768	29	CC008801
c 40	812	9	AL695087
c 41	815	29	CC008799
c 42	816	28	AY079596
c 43	911	13	BU090243
c 44	928	29	BZ567557
c 45	1085	29	CNS077DK

#### ALIGNMENTS

#### RESULT 1

BZ568959/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BZ568959 1445 bp DNA linear GSS 17-DEC-2002  
pacs2-164\_8171.y2 pacs2-164 Pseudomonas aeruginosa genomic clone  
pacs2-164\_8171, genomic survey sequence.

BZ568959 GI:27202799

GSS.

Pseudomonas aeruginosa

Pseudomonas aeruginosa

Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;

Pseudomonadaceae; Pseudomonas.

1 (bases 1 to 1445)

Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,

Burns, J.L., Kaul, R. and Olsen, M.V.

Whole-Genome-Sequence variation among multiple isolates of

Pseudomonas aeruginosa library

J. Bacteriol., (2002) In press

Contact: Chris K. Raymond

Genome Center

University of Washington

Box 352145, Seattle, WA 98105-2145, USA

Tel.: 2062216954

Fax: 2066857244

Email: craymond@u.washington.edu

Class: shotgun.

Location/Qualifiers

```

source
1. 1445
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="2-164"
/db_xref="taxon:287"
/clone="pac2-164_8171"
/clone_lib="pac2-164"
/notes="clinical isolate 2-164 Whole genomic shotgun library."
BASE COUNT 308 a 424 c 412 g 293 t 8 others
ORIGIN
Query Match 41.4%; Score 600; DB 29; Length 1445;
Best Local Similarity 100.0%; Pred. No. 9.3e-289;
Matches 600; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 488 CCGGGGCTCGACCTGCTGGCCGACGACATCGAGCTGTAGCTGTGAGCTGTCGAGCTT 547
Db 676 CCGGGGCTCGACCTGCTGGCCGACGACATCGAGCTGTAGCTGTGAGCTGTCGAGCTT 617
QY 548 CCAGCTGGAACCTGGGATCGCTCAACGCCGAGGTGGCGACCGCTGCTGAACCTCAGCGA 607
Db 616 CCAGCTGGAACCTGGGATCGCTCAACGCCGAGGTGGCGACCGCTGCTGAACCTCAGCGA 557
QY 608 AGACCATATGATCGCTACAGCGATGGCTGACTACCACTGCGCAAGCAACCGATCTT 667
Db 556 AGACCATATGATCGCTACAGCGATGGCTGACTACCACTGCGCAAGCAACCGATCTT 497
QY 668 CCGCGTGGCCGCGAGTGTGTGAATCGCGCGGATGCCCTGACCCGACCGCTGATCGC 727
Db 496 CCGCGTGGCCGCGAGTGTGTGAATCGCGCGGATGCCCTGACCCGACCGCTGATCGC 437
QY 728 CGATACCGTGGCTGTGCTTCCGCTTGAACAGCGGACTTCAAGGCTTTCCGGCTT 787
Db 436 CGATACCGTGGCTGTGCTTCCGCTTGAACAGCGGACTTCAAGGCTTTCCGGCTT 377
QY 788 GATCGAGGAGAGCGGCCAGAGTGGCTGGCTTCCAGTTCGACAACTGTCGCCGCTTG 847
Db 376 GATCGAGGAGAGCGGCCAGAGTGGCTTCCAGTTCGACAACTGTCGCCGCTTG 317
QY 848 CGAAGTGAAGTCCGTGGCGCCACAACTATTCGAACGGCTGCGCGCTGCGCTGG 907
Db 316 CGAAGTGAAGTCCGTGGCGCCACAACTATTCGAACGGCTGCGCGCTGCGCTGG 257
QY 908 CCATCGGCTGGCTGCGCTTGCACGCCATGCTCGCGCGCTGGAAGCGCTTTTCCGGCT 967
Db 256 CCATCGGCTGGCTGCGCTTGCACGCCATGCTCGCGCGCTGGAAGCGCTTTTCCGGCT 197
QY 968 GGCATATCGCTGCGCTGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1027
Db 196 GGCATATCGCTGCGCTGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 137
QY 1028 GGCACCAAGTGGCGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1087
Db 136 GGCACCAAGTGGCGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 77

```

```

RESULT 2
Bz579116/c
LOCUS
DEFINITION msh2_6205.y2 msh Pseudomonas aeruginosa genomic clone msh2_6205,
genomic survey sequence.
ACCESSION Bz579116
VERSION Bz579116.1 GI:27214177
KEYWORDS GSS.
SOURCE Pseudomonas aeruginosa
ORGANISM Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE 1 (bases 1 to 892)
AUTHORS Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE Whole-Genome-Sequence variation among multiple isolates of

```

```

Pseudomonas aeruginosa library
J. Bacteriol., (2002) In press
Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.
Location/Qualifiers
1. 892
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="MSH"
/db_xref="taxon:287"
/clone="msh2_6205"
/clone_lib="msh"
/notes="Environmental isolate. Whole genomic shotgun library."
BASE COUNT 173 a 277 c 276 g 160 t 6 others
ORIGIN
Query Match 35.2%; Score 510; DB 29; Length 892;
Best Local Similarity 99.8%; Pred. No. 8.4e-244;
Matches 560; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 319 CCGGAAAGCGGTGGCGATCTCGGGTGACATCGATCTTTCGCCCGCGAGCGGAAGGCC 378
Db 644 CCGGAAAGCGGTGGCGATCTCGGGTGACATCGATCTTTCGCCCGCGAGCGGAAGGCC 585
QY 379 CGATCTGCGCCATCACCGGTTCACGCGAAGACGACCGTGCACCCCTGGTGGCGGAAA 438
Db 584 CGATCTGCGCCATCACCGGTTCACGCGAAGACGACCGTGCACCCCTGGTGGCGGAAA 525
QY 439 TGGCGGTGGCGGGGACAAAGCGTGTCCCGTGGCGGCAACCTCGCGACCCCGCGCTCG 498
Db 524 TGGCGGTGGCGGGGACAAAGCGTGTCCCGTGGCGGCAACCTCGCGACCCCGCGCTCG 465
QY 499 ACTGCTGCGCGACGACATCGAGCTGTAGCTGTGGAGCTGTGAGCTTCCAGCTGAAA 558
Db 464 ACTGCTGCGCGACGACATCGAGCTGTAGCTGTGGAGCTGTGAGCTTCCAGCTGAAA 405
QY 559 CTTGCGATCCCTCAACGCCGAGGTGGCGACCGTGTGAACGTCAGCGAAGACCATATGG 618
Db 404 CTTGCGATCCCTCAACGCCGAGGTGGCGACCGTGTGAACGTCAGCGAAGACCATATGG 345
QY 619 ATCGCTACGACGCGATGGCTGACTACCACTGGCCCAAGCACCGGATCTTCCGGGTGCC 678
Db 344 ATCGCTACGACGCGATGGCTGACTACCACTGGCCCAAGCACCGGATCTTCCGGGTGCC 285
QY 679 GCGAGTCTGTGTGAATCGCGCGCGATGCCCTGACCGACCGCTGATCGCGATACCGTGC 738
Db 284 GCGAGTCTGTGTGAATCGCGCGCGATGCCCTGACCGACCGCTGATCGCGATACCGTGC 225
QY 739 CCGTGTGTGCTGTGCGGCTTGAACAGCCGAGCTTCAAGGCTTTCGGCCCTGATCGAAG 798
Db 224 CCGTGTGTGCTGTGCGGCTTGAACAGCCGAGCTTCAAGGCTTTCGGCCCTGATCGAAG 165
QY 799 ACGGCGAAGTGGCTGGGTTCCAGTTCCAGAACTGCTGCCGTTGGCGAAGTGAAGA 858
Db 164 ACGGCGAAGTGGCTGGGTTCCAGTTCCAGAACTGCTGCCGTTGGCGAAGTGAAGA 105
QY 859 TCCGTGGCGGCCACCAACTATT 879
Db 104 TCCGTGGCGGCCACCAACTATT 84

```

```

RESULT 3
Bz559806
LOCUS
DEFINITION Bz559806
1171 bp DNA linear GSS 17-DEC-2002
pac2-164_1863.xl pac2-164 Pseudomonas aeruginosa genomic clone
pac2-164_1863, genomic survey sequence.

```



B2550283  
 B2550283.1 GI:27153864  
 GSS.  
 Pseudomonas aeruginosa  
 Pseudomonas aeruginosa  
 Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
 Pseudomonadaceae; Pseudomonas.  
 1 (bases 1 to 1194)  
 Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,  
 Burns, J.L., Kaul, R. and Olsen, M.V.  
 Whole-Genome-Sequence variation among multiple isolates of  
 Pseudomonas aeruginosa library  
 J. Bacteriol., (2002) In press  
 Contact: Chris K. Raymond  
 Genome Center  
 University of Washington  
 Box 352145, Seattle, WA 98105-2145, USA

ORIGIN

ORIGIN

```

Query Match      1.6%; Score 23; DB 10; Length 1754;
Best Local Similarity 100.0%; Pred. No. 4.6;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 884 CGCGCTCGCGCGCTGCGCTGG 906
      |||||
Db 343 CGCGCTCGCGCGCTGCGCTGG 321

RESULT 6
AL825430/c
LOCUS AL825430 p:335 Triticum aestivum cDNA clone B08_p335_plate_15, mRNA
DEFINITION sequence.
ACCESSION AL825430
VERSION AL825430.1 GI:21836951
KEYWORDS EST.
SOURCE Triticum aestivum (bread wheat)
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
; Triticeae; Triticum.
REFERENCE 1 (bases 1 to 537)
AUTHORS Wilton,I., Beswick,R., Shepherd,S., Barker,G., Parker,J., Owen,P.,
Edwards,D., Coghill,J., Holdsworth,M., Lenton,J., Shewry,P. and
Edwards,K.
TITLE A BBSRC-funded wheat EST resource for the academic community
JOURNAL Unpublished
COMMENT Contact: Barker G
Institute of Arable Crop Research
Long Ashton, Bristol BS41 9AF United Kingdom.
FEATURES
source
1. .537
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="mercia"
/db_xref="taxon:4565"
/clone="B08_p335_plate_15"
/tissue_type="salt stressed seedlings"
/dev_stage="21 days old"
/clone_lib="P:335"
BASE COUNT 86 a 216 c 147 g 87 t 1 others
ORIGIN

Query Match      1.5%; Score 22; DB 9; Length 537;
Best Local Similarity 100.0%; Pred. No. 9.9;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1274 CGCGAGCTGCGCGCGAGGC 1295
      |||||
Db 117 CGCGAGCTGCGCGCGAGGC 96

RESULT 7
BE586183/c
LOCUS BE586183 p:687 K5U wheat Fusarium graminearum infected
DEFINITION spike cDNA library Triticum aestivum cDNA clone
EST7-4-5_C02_7elt7_006, mRNA sequence.
ACCESSION BE586183
VERSION BE586183.1 GI:9839203
KEYWORDS EST.
SOURCE Triticum aestivum (bread wheat)
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
; Triticeae; Triticum.
REFERENCE 1 (bases 1 to 687)
AUTHORS Fellers,J.P., Li,W.L., Hill-Ambroz,K., Matthews,A. and Gill,B.S.
TITLE The structure and function of the expressed portion of the wheat
genomes - Kansas State University. Fusarium graminearum infected
spike cDNA library
JOURNAL Unpublished

```

Contact: John Fellers  
 US Department of Agriculture, Agriculture Research Service, Plant  
 Science and Entomology Unit  
 Dept. of Plant Pathology, 4006 Throckmorton Hall, Kansas State  
 University, Manhattan, KS 66506, USA  
 Tel: 785-532-2367  
 Fax: 785-532-6167  
 Email: jpf@alfalfa.ksu.edu  
 Sequence have been trimmed to remove vector sequence and low  
 quality sequence with phred score less than 20  
 Seq primer: 17.

FEATURES  
 source  
 1. .687  
 /organism="Triticum aestivum"  
 /mol\_type="mRNA"  
 /cultivar="Sumai3"  
 /db\_xref="taxon:4565"  
 /clone="Est7-4-5\_C02\_7elt7\_006"  
 /tissue\_type="Spike"  
 /dev\_stage="Adult plant"  
 /lab\_host="E. coli JM109"  
 /clone\_lib="KSU wheat Fusarium graminearum infected spike  
 cDNA library"  
 /note="Vector: pGEM-T easy; Site\_1: SacII; Site\_2: SpeI;  
 Plants were grown in the greenhouse. Spikes were sprayed  
 with Fusarium graminearum (at what stage). Total RNA, and  
 poly(A) RNA were prepared from infected spikes. cDNA was  
 prepared using the SmartW PCR cDNA synthesis kit from  
 Clontech. cDNA was cloned into the pGEM-T easy vector  
 from Promega."

BASE COUNT 140 a 246 c 182 g 117 t 2 others  
 ORIGIN

Query Match 1.5%; Score 22; DB 10; Length 687;  
 Best Local Similarity 100.0%; Pred. No. 11;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1274 CGCGAGCTGCGCGCGAGGC 1295  
 |||||  
 Db 428 CGCGAGCTGCGCGCGAGGC 407

RESULT 8  
 CB061196/c  
 LOCUS CB061196 265 bp mRNA linear EST 17-JAN-2003  
 DEFINITION 4011803 BARC-EMBRAPA 338BOV Bos indicus cDNA clone 338BOV\_1C19  
 Unknown, mRNA sequence.  
 ACCESSION CB061196  
 VERSION CB061196.1 GI:27799483  
 KEYWORDS EST.  
 SOURCE Bos indicus (zebu)  
 ORGANISM Bos indicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;  
 Bovidae; Bovinae; Bos.  
 REFERENCE 1 (bases 1 to 265)  
 AUTHORS da Mota,A.F., Sonstegard,T.S., Van Tassel,C.P., Matukumalli,L.K.,  
 Wood,D.L., Capuco,A.V., Brito,M.A.P., Martinez,M.L., Connor,E.E.,  
 Machado,M.A. and Coutinho,L.L.  
 TITLE Construction and Characterization of cDNA Libraries Generated from  
 Mammary Gland Tissues of Holstein (Bos taurus) and Gir (Bos indicus  
 ) Cattle  
 JOURNAL Unpublished  
 COMMENT Contact: Adilson F. da Mota  
 Gene Evaluation and Mapping Laboratory  
 USDA, ARS, Animal and Natural Resources Institute  
 Bldg. 200 Rm3 BARC-East, Beltsville, MD 20705, USA  
 Tel: 3015048456  
 Fax: 3015048414  
 Email: amota@npgi.embrapa.br  
 Single pass sequencing. Bases called and trimmed with phred  
 0.000925 using options '-trim\_alt'' -trim\_fasta. Vector identified  
 by cross\_match using options -minmatch 12 -minscore 12

PCR Primers

FORWARD: GTTTCACGTCACGAGTTG  
 BACKWARD: TGAGCGGATCAATTTACACAG  
 Plate: 1 row: C column: 19  
 Seq primer: GTTTCACGTCACGAGTTG  
 High quality sequence stop: 265.  
 Location/Qualifiers

FEATURES

source  
 1..265  
 /organism="Bos indicus"  
 /mol\_type="mRNA"  
 /strain="Brazilian Dairy Gir"  
 /db\_xref="taxon:9915"  
 /clone="338BOV\_1C19"  
 /sex="female"  
 /tissue\_type="teat cistern and alveoli"  
 /cell\_type="multiple"  
 /dev\_stage="involted"  
 /lab\_host="K-12"  
 /clone\_lib="BARC-EMBRAPA 338BOV"  
 /note="Organ: mammary; Vector: pUC 118; Site\_1: HincII;  
 Site\_2: HincII; This mammary-derived cDNA library was  
 created as part of a collaborative project between the ARS  
 Gene Evaluation and Mapping Laboratory and the EMBRAPA  
 Dairy Cattle Research Center under the sponsorship of USDA  
 , ARS/EMBRAPA-LABEX program in animal genomics. RNA  
 extracted on 6/8/02, RT with Superscript II at 37 deg C  
 annealing temperature. Cow AM1." 46 t

BASE COUNT 52 a 75 c 92 g 46 t  
 ORIGIN  
 Query Match 1.4%; Score 21; DB 14; Length 265;  
 Best Local Similarity 100.0%; Pred. No. 25;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 900 GCGCTGGCCATCGCGTGGC 920  
 Db 255 GCGCTGGCCATCGCGTGGC 235

RESULT 9

AA969049 369 bp mRNA linear EST 07-JUL-1998  
 LOCUS op43e08.sl Soares\_NFL\_T\_GBC\_S1 Homo sapiens cDNA clone  
 DEFINITION IMAGE:1579622 3', mRNA sequence.

ACCESSION AA969049  
 VERSION AA969049.1 GI:3144229

KEYWORDS EST.  
 SOURCE Homo sapiens (human)

ORGANISM  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 369)  
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index

JOURNAL Unpublished  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-r@mail.nih.gov

This clone is available royalty-free through LNL ; contact the  
 IMAGE Consortium (info@image.lnl.gov) for further information.  
 Insert Length: 630 Std Error: 0.00  
 Seq primer: -40m13 fwd. Et from Amersham  
 High quality sequence stop: 334.  
 Location/Qualifiers

FEATURES

source  
 1..369  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:1579622"  
 /lab\_host="DH10B"  
 /clone\_lib="Soares\_NFL\_T\_GBC\_S1"  
 /note="Organ: pooled; Vector: pT7m3D-Pac (Pharmacia) with  
 a modified polylinker; Site\_1: Not I; Site\_2: Eco RI;

Equal amounts of plasmid DNA from three normalized  
 libraries (fetal lung NBHL19W, testis NHT, and B-cell  
 NCI-CGAP-GCB1) were mixed, and ss circles were made in  
 vitro. Following HAP purification, this DNA was used as  
 tracer in a subtractive hybridization reaction. The driver  
 was PCR-amplified cDNAs from pools of 5,000 clones made  
 from the same 3 libraries. The pools consisted of  
 I.M.A.G.E. clones 297480-302087, 682632-687239,  
 726408-728711, and 729096-731399. Subtraction by Bento  
 Soares and M. Fatima Bonaldo.

BASE COUNT 85 a 87 c 107 g 90 t  
 ORIGIN

Query Match 1.4%; Score 21; DB 9; Length 369;  
 Best Local Similarity 100.0%; Pred. No. 28;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1268 GCAGCGCCGCGAGTGGCCCG 1288  
 Db 124 GCAGCGCCGCGAGTGGCCCG 144

RESULT 10

BI306480/c  
 LOCUS

DEFINITION NL\_4\_N22 Drought stress (leaf) Oryza sativa (indica cultivar-group)  
 CDNA clone NL\_4\_N22 3', mRNA sequence.

ACCESSION BI306480  
 VERSION BI306480.1 GI:14981802

KEYWORDS EST.

SOURCE Oryza sativa (indica cultivar-group)  
 ORGANISM Oryza sativa (indica cultivar-group)

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE 1 (bases 1 to 501)  
 AUTHORS Reddy,A.R., Ramakrishna,W., Chandrasekhar,A., Nagabhushan,I.,  
 Ravindrababu,P. and Bennetzen,J.L.

TITLE Novel EST enrichment with normalized cDNA libraries from drought  
 stressed rice (Oryza sativa L.cv Nagina 22)

JOURNAL Unpublished  
 COMMENT Contact: Reddy AR

Department of Plant Sciences, School of Life Sciences  
 University of Hyderabad

P.O. Central University, Hyderabad-500 046, A.P, India  
 Tel.: 0091-40-3010265

Fax: 0091-40-3010145  
 Email: arjuls@uohyd.ernet.in

Insert Length: 501 Std Error: 0.00  
 Plate: 4 row: N column: 22

Seq primer: GTAAACCGCGCCAGTG.  
 Location/Qualifiers

source  
 1..501

/organism="Oryza sativa (indica cultivar-group)"  
 /mol\_type="mRNA"  
 /cultivar="Nagina 22 (indica sub sp)"  
 /db\_xref="taxon:39946"  
 /clone="NL\_4\_N22"

/tissue\_type="Entire leaf tissue"  
 /dev\_stage="35 day-old seedlings"  
 /clone\_lib="Drought stress (leaf)"

/note="Organ: Leaf; Vector: T7m3Pac; ESTs from normalized  
 leaf cDNA Library from drought stressed seedlings"  
 BASE COUNT 52 a 140 c 226 g 82 t 1 others

ORIGIN

Query Match 1.4%; Score 21; DB 12; Length 501;  
 Best Local Similarity 100.0%; Pred. No. 31;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1050 CTGGCGCGATCGAGGGCTG 1070  
 Db 415 CTGGCGCGATCGAGGGCTG 395

DNA sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL, send email to:  
 info@image.llnl.gov  
 Seq primer: -400P from Gibco  
 High quality sequence stop: 268.

RESULT 11  
 BE041029  
 LOCUS  
 DEFINITION OF16F05 OF Oryza sativa cDNA 5' similar to histone-like dna-binding protein pf 1, mRNA sequence.  
 ACCESSION BE041029  
 VERSION  
 KEYWORDS  
 SOURCE EST.  
 ORGANISM Oryza sativa

1. .286  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:3032243"  
 /lab\_host="DH10B"  
 /clone\_lib="NCI-CGAP\_Kid11"  
 /note="Organ: kidney; Vector: pMT3D-Pac (Pharmacia) with a modified polylinker; Site: 1; Not 1; Site: 2; Eco RI; Plasmid DNA from the normalized library NCI-CGAP\_Kid3 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIDs 1322376-1323911, 1456007-1456775, and 1500552-1502855). Subtraction by Bento Soares and M. Fatima Bonaldo."

REFERENCE  
 AUTHORS Bohnert,H.J., Borchert,C., Brazille,S., Brooks,J., Eaton,M., Ferreira,H., Kawasaki,S., McColough,A., Michalowski,C.B., Palacio,C., Scara,G., Wheeler,M. and Zepeda,G.R.  
 TITLE Functional Genomics of Plant Stress Tolerance  
 JOURNAL Unpublished  
 COMMENT Contact: Michalowski,C.B.  
 University of Arizona  
 Bio Sciences West room 513, Tucson, AZ 85721, USA  
 Tel: 520-621-7982  
 Fax: 520-621-1697  
 Email: cbm@u.arizona.edu  
 An open reading frame exists.

FEATURES  
 source  
 1. .792  
 Location/Qualifiers  
 /organism="Oryza sativa"  
 /mol\_type="mRNA"  
 /strain="pokkali"  
 /db\_xref="taxon:4530"  
 /tissue\_type="entire plant"  
 /dev\_stage="2 weeks"  
 /clone\_lib="OF"  
 /note="11 week 150mM NaCl"

BASE COUNT 77 a 189 c 181 g 345 t  
 ORIGIN

Query Match 1.4%; Score 21; DB 10; Length 792;  
 Best Local Similarity 100.0%; Pred. No. 36;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1050 CTGGCGCGCATCGAGGGCTG 1070  
 ||||||||||||||||||  
 Db 54 CTGGCGCGCATCGAGGGCTG 74

RESULT 12  
 AW771204  
 LOCUS  
 DEFINITION hn60b06.x1 NCI-CGAP\_Kid11 Homo sapiens cDNA clone IMAGE:3032243 3', mRNA sequence.  
 ACCESSION AW771204  
 VERSION  
 KEYWORDS  
 SOURCE EST.  
 ORGANISM Homo sapiens (human)

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-remail.nih.gov  
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.  
 CDNA Library Preparation: M. Bento Soares, Ph.D.  
 CDNA Library Arrayed by: Greg Lennon, Ph.D.

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-remail.nih.gov  
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.  
 CDNA Library Preparation: M. Bento Soares, Ph.D.  
 CDNA Library Arrayed by: Greg Lennon, Ph.D.

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-remail.nih.gov  
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.  
 CDNA Library Preparation: M. Bento Soares, Ph.D.  
 CDNA Library Arrayed by: Greg Lennon, Ph.D.

CDNA Library Arrayed by: Greg Lennon, Ph.D.

# FEATURES

## source

1. .286  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:3032243"  
 /lab\_host="DH10B"  
 /clone\_lib="NCI-CGAP\_Kid11"  
 /note="Organ: kidney; Vector: pMT3D-Pac (Pharmacia) with a modified polylinker; Site: 1; Not 1; Site: 2; Eco RI; Plasmid DNA from the normalized library NCI-CGAP\_Kid3 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIDs 1322376-1323911, 1456007-1456775, and 1500552-1502855). Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 76 a 56 c 71 g 83 t  
 ORIGIN

Query Match 1.4%; Score 20; DB 9; Length 286;  
 Best Local Similarity 100.0%; Pred. No. 80;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1268 GCAGCGCGCGAGCTGCC 1287  
 ||||||||||||||||||  
 Db 147 GCAGCGCGCGAGCTGCC 166

RESULT 13  
 CB042199  
 LOCUS

DEFINITION 4007144 BARC-EMBRAPA 316BOV Bos indicus cDNA clone 316BOV\_1G12  
 Unknown, mRNA sequence.

ACCESSION CB042199  
 VERSION  
 KEYWORDS  
 SOURCE EST.  
 ORGANISM Bos indicus (zebu)

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT

Contact: Adilson F. da Mota  
 Gene Evaluation and Mapping Laboratory  
 USDA, ARS, Animal and Natural Resources Institute  
 Bldg. 200 Rm3 BARC-East, Beltsville, MD 20705, USA  
 Tel: 3015048456  
 Fax: 3015048414  
 Email: amota@cnpl.embrapa.br  
 Single pass sequencing. Bases called and trimmed with phred  
 0.000925 using options -trim\_alt -trim\_fasta. Vector identified  
 by cross\_match using options -minmatch 12 -minscore 18  
 PCR Primers  
 FORWARD: GTTTCCTCCAGTCACGCGTTG  
 BACKWARD: TGAGCGGATACAAATTCACACAG  
 Plate: 1 row: G column: 12  
 Seq primer: GTTTCCTCCAGTCACGCGTTG  
 High quality sequence stop: 289.

CB042199  
 LOCUS

DEFINITION 4007144 BARC-EMBRAPA 316BOV Bos indicus cDNA clone 316BOV\_1G12  
 Unknown, mRNA sequence.

ACCESSION CB042199  
 VERSION  
 KEYWORDS  
 SOURCE EST.  
 ORGANISM Bos indicus (zebu)

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT

Contact: Adilson F. da Mota  
 Gene Evaluation and Mapping Laboratory  
 USDA, ARS, Animal and Natural Resources Institute  
 Bldg. 200 Rm3 BARC-East, Beltsville, MD 20705, USA  
 Tel: 3015048456  
 Fax: 3015048414  
 Email: amota@cnpl.embrapa.br  
 Single pass sequencing. Bases called and trimmed with phred  
 0.000925 using options -trim\_alt -trim\_fasta. Vector identified  
 by cross\_match using options -minmatch 12 -minscore 18  
 PCR Primers  
 FORWARD: GTTTCCTCCAGTCACGCGTTG  
 BACKWARD: TGAGCGGATACAAATTCACACAG  
 Plate: 1 row: G column: 12  
 Seq primer: GTTTCCTCCAGTCACGCGTTG  
 High quality sequence stop: 289.

CB042199  
 LOCUS

DEFINITION 4007144 BARC-EMBRAPA 316BOV Bos indicus cDNA clone 316BOV\_1G12  
 Unknown, mRNA sequence.

ACCESSION CB042199  
 VERSION  
 KEYWORDS  
 SOURCE EST.  
 ORGANISM Bos indicus (zebu)

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT

Contact: Adilson F. da Mota  
 Gene Evaluation and Mapping Laboratory  
 USDA, ARS, Animal and Natural Resources Institute  
 Bldg. 200 Rm3 BARC-East, Beltsville, MD 20705, USA  
 Tel: 3015048456  
 Fax: 3015048414  
 Email: amota@cnpl.embrapa.br  
 Single pass sequencing. Bases called and trimmed with phred  
 0.000925 using options -trim\_alt -trim\_fasta. Vector identified  
 by cross\_match using options -minmatch 12 -minscore 18  
 PCR Primers  
 FORWARD: GTTTCCTCCAGTCACGCGTTG  
 BACKWARD: TGAGCGGATACAAATTCACACAG  
 Plate: 1 row: G column: 12  
 Seq primer: GTTTCCTCCAGTCACGCGTTG  
 High quality sequence stop: 289.

CB042199  
 LOCUS

DEFINITION 4007144 BARC-EMBRAPA 316BOV Bos indicus cDNA clone 316BOV\_1G12  
 Unknown, mRNA sequence.

ACCESSION CB042199  
 VERSION  
 KEYWORDS  
 SOURCE EST.  
 ORGANISM Bos indicus (zebu)

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT

Contact: Adilson F. da Mota  
 Gene Evaluation and Mapping Laboratory  
 USDA, ARS, Animal and Natural Resources Institute  
 Bldg. 200 Rm3 BARC-East, Beltsville, MD 20705, USA  
 Tel: 3015048456  
 Fax: 3015048414  
 Email: amota@cnpl.embrapa.br  
 Single pass sequencing. Bases called and trimmed with phred  
 0.000925 using options -trim\_alt -trim\_fasta. Vector identified  
 by cross\_match using options -minmatch 12 -minscore 18  
 PCR Primers  
 FORWARD: GTTTCCTCCAGTCACGCGTTG  
 BACKWARD: TGAGCGGATACAAATTCACACAG  
 Plate: 1 row: G column: 12  
 Seq primer: GTTTCCTCCAGTCACGCGTTG  
 High quality sequence stop: 289.

CB042199  
 LOCUS

DEFINITION 4007144 BARC-EMBRAPA 316BOV Bos indicus cDNA clone 316BOV\_1G12  
 Unknown, mRNA sequence.

ACCESSION CB042199  
 VERSION  
 KEYWORDS  
 SOURCE EST.  
 ORGANISM Bos indicus (zebu)

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT

Contact: Adilson F. da Mota  
 Gene Evaluation and Mapping Laboratory  
 USDA, ARS, Animal and Natural Resources Institute  
 Bldg. 200 Rm3 BARC-East, Beltsville, MD 20705, USA  
 Tel: 3015048456  
 Fax: 3015048414  
 Email: amota@cnpl.embrapa.br  
 Single pass sequencing. Bases called and trimmed with phred  
 0.000925 using options -trim\_alt -trim\_fasta. Vector identified  
 by cross\_match using options -minmatch 12 -minscore 18  
 PCR Primers  
 FORWARD: GTTTCCTCCAGTCACGCGTTG  
 BACKWARD: TGAGCGGATACAAATTCACACAG  
 Plate: 1 row: G column: 12  
 Seq primer: GTTTCCTCCAGTCACGCGTTG  
 High quality sequence stop: 289.

CB042199  
 LOCUS

DEFINITION 4007144 BARC-EMBRAPA 316BOV Bos indicus cDNA clone 316BOV\_1G12  
 Unknown, mRNA sequence.

ACCESSION CB042199  
 VERSION  
 KEYWORDS  
 SOURCE EST.  
 ORGANISM Bos indicus (zebu)

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT

Contact: Adilson F. da Mota  
 Gene Evaluation and Mapping Laboratory  
 USDA, ARS, Animal and Natural Resources Institute  
 Bldg. 200 Rm3 BARC-East, Beltsville, MD 20705, USA  
 Tel: 3015048456  
 Fax: 3015048414  
 Email: amota@cnpl.embrapa.br  
 Single pass sequencing. Bases called and trimmed with phred  
 0.000925 using options -trim\_alt -trim\_fasta. Vector identified  
 by cross\_match using options -minmatch 12 -minscore 18  
 PCR Primers  
 FORWARD: GTTTCCTCCAGTCACGCGTTG  
 BACKWARD: TGAGCGGATACAAATTCACACAG  
 Plate: 1 row: G column: 12  
 Seq primer: GTTTCCTCCAGTCACGCGTTG  
 High quality sequence stop: 289.

CB042199  
 LOCUS

DEFINITION 4007144 BARC-EMBRAPA 316BOV Bos indicus cDNA clone 316BOV\_1G12  
 Unknown, mRNA sequence.

ACCESSION CB042199  
 VERSION  
 KEYWORDS  
 SOURCE EST.  
 ORGANISM Bos indicus (zebu)

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT

Contact: Adilson F. da Mota  
 Gene Evaluation and Mapping Laboratory  
 USDA, ARS, Animal and Natural Resources Institute  
 Bldg. 200 Rm3 BARC-East, Beltsville, MD 20705, USA  
 Tel: 3015048456  
 Fax: 3015048414  
 Email: amota@cnpl.embrapa.br  
 Single pass sequencing. Bases called and trimmed with phred  
 0.000925 using options -trim\_alt -trim\_fasta. Vector identified  
 by cross\_match using options -minmatch 12 -minscore 18  
 PCR Primers  
 FORWARD: GTTTCCTCCAGTCACGCGTTG  
 BACKWARD: TGAGCGGATACAAATTCACACAG  
 Plate: 1 row: G column: 12  
 Seq primer: GTTTCCTCCAGTCACGCGTTG  
 High quality sequence stop: 289.

CB042199  
 LOCUS

DEFINITION 4007144 BARC-EMBRAPA 316BOV Bos indicus cDNA clone 316BOV\_1G12  
 Unknown, mRNA sequence.

ACCESSION CB042199  
 VERSION  
 KEYWORDS  
 SOURCE EST.  
 ORGANISM Bos indicus (zebu)

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT

Contact: Adilson F. da Mota  
 Gene Evaluation and Mapping Laboratory  
 USDA, ARS, Animal and Natural Resources Institute  
 Bldg. 200 Rm3 BARC-East, Beltsville, MD 20705, USA  
 Tel: 3015048456  
 Fax: 3015048414  
 Email: amota@cnpl.embrapa.br  
 Single pass sequencing. Bases called and trimmed with phred  
 0.000925 using options -trim\_alt -trim\_fasta. Vector identified  
 by cross\_match using options -minmatch 12 -minscore 18  
 PCR Primers  
 FORWARD: GTTTCCTCCAGTCACGCGTTG  
 BACKWARD: TGAGCGGATACAAATTCACACAG  
 Plate: 1 row: G column: 12  
 Seq primer: GTTTCCTCCAGTCACGCGTTG  
 High quality sequence stop: 289.

CB042199  
 LOCUS

DEFINITION 4007144 BARC-EMBRAPA 316BOV Bos indicus cDNA clone 316BOV\_1G12  
 Unknown, mRNA sequence.

ACCESSION CB042199  
 VERSION  
 KEYWORDS  
 SOURCE EST.  
 ORGANISM Bos indicus (zebu)

FEATURES  
sourceLocation/Qualifiers  
1. .289

/organism="Bos indicus"  
/mol\_type="mRNA"  
/strain="Brazilian Dairy Gir"  
/db\_xref="taxon:9915"  
/clone="316BOV\_1G12"  
/sex="female"  
/tissue\_type="mammary"  
/cell\_type="epithelium"  
/dev\_stage="involved"  
/lab\_host="K-12"  
/clone\_lib="BARC-EMBRAPA 316BOV"  
/note="Organ: mammary; Vector: pUC19; Site\_1: Smal;  
Site\_2: SmaI; This mammary-derived cDNA library was  
created as part of a collaborative project between the ARS  
Gene Evaluation and Mapping Laboratory and the EMBRAPA  
Dairy Cattle Research Center under the sponsorship of USDA  
, ARS/EMBRAPA-LABEX program in animal genomics. RNA  
extracted on 6/8/02, RT with Superscript II at 37 deg C,  
annealing temperature, PCR with 16mer AM5."

BASE COUNT  
ORIGIN

47 a 100 c 87 g 55 t

Query Match 1.4%; Score 20; DB 14; Length 289;  
Best Local Similarity 100.0%; Pred. No. 81;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  
Db

901 CGCTGGGCGCATCGGTCGCGC 920  
|||||  
1 CGCTGGGCCATCGGTCGCGC 20

RESULT 14  
BF657011/c

LOCUS BF657011 301 bp mRNA linear EST 20-DEC-2000  
DEFINITION OV2\_18.A07.g1.A002 Ovary 2 (OV2) Sorghum bicolor cDNA, mRNA  
sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE

BF657011  
BF657011.1 GI:11922145  
EST.  
Sorghum bicolor (sorghum)

## ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Sorghum.

REFERENCE  
AUTHORS

1 (bases 1 to 301)  
Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt  
L.H.

TITLE  
JOURNAL  
COMMENT

An EST database from Sorghum: ovaries of varying immature stages  
Unpublished  
Contact: Cordonnier-Pratt MM  
Laboratory for Genomics and Bioinformatics  
The University of Georgia, Department of Plant Biology  
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
Tel: 706 542 1860  
Fax: 706 583 0210  
Email: mmpratt@uga.edu  
Seq primer: T7  
High quality sequence start: 6  
High quality sequence stop: 222  
POLYA=Yes.

FEATURES  
sourceLocation/Qualifiers  
1. .301

/organism="Sorghum bicolor"  
/mol\_type="mRNA"  
/db\_xref="taxon:4558"  
/clone\_lib="Ovary 2 (OV2)"  
/note="Organ: Mix of ovaries of varying immature stages  
from 8-week-old plants; Vector: pBluescript II from Lambda  
Zap II; Site\_1: XhoI; Site\_2: EcoRI; The library was made  
from poly-A RNA in the cloning vector lambda ZAP II.  
Clones to be sequenced were prepared by mass excision."

BASE COUNT  
ORIGIN

76 a 62 c 83 g 80 t

## Query Match

1.4%; Score 20; DB 10; Length 301;

## Best Local Similarity

100.0%; Pred. No. 82;

## Matches

20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  
Db

1305 CTGTTGTGCGCGGCTGCGC 1324  
|||||  
37 CTGTTGTGCGCGGCTGCGC 18

RESULT 15  
AW139275

LOCUS AW139275 318 bp mRNA linear EST 30-OCT-1999  
DEFINITION UI-H-B11-aeg-g-12-0-UI.s1 NCI\_CGAP\_Sub3 Homo sapiens cDNA clone  
IMAGE:2719318 3', mRNA sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE

AW139275  
AW139275.1 GI:6143993  
EST.  
Homo sapiens (human)

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Unpublished  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Oligo-dT track not found, Not I site shown in beginning of sequence  
is likely internal to the message. cDNA Library Preparation: M.B.  
Soares Lab Clone distribution: NCI-CGAP clone distribution  
information can be found through the I.M.A.G.E. Consortium/LLNL at:  
www-bio.llnl.gov/bbrp/image/image.html  
Seq primer: M13 Forward  
POLYA=No.

FEATURES  
sourceLocation/Qualifiers  
1. .318

/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2719318"  
/lab\_host="DH10B (Life Technologies)"  
/clone\_lib="NCI\_CGAP\_Sub3"  
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: Eco RI; The  
NCI\_CGAP\_Sub3 library is a subtracted library derived from  
the NCI\_CGAP\_Sub1 library, which is a subtracted library  
derived from BI. BI constitutes a mixture of 21  
normalized or subtracted NCI\_CGAP libraries: NCI\_CGAP\_C04  
, NCI\_CGAP\_Pr22, NCI\_CGAP\_Pr28, NCI\_CGAP\_C010,  
NCI\_CGAP\_C016, NCI\_CGAP\_Kid5, NCI\_CGAP\_Kid12,  
NCI\_CGAP\_Kid3, NCI\_CGAP\_Kid11, NCI\_CGAP\_Lym2,  
NCI\_CGAP\_Br2, NCI\_CGAP\_C08, NCI\_CGAP\_C11, NCI\_CGAP\_Lei2,  
NCI\_CGAP\_Brn23, NCI\_CGAP\_Lu5, NCI\_CGAP\_Lu24,  
NCI\_CGAP\_Lu19, NCI\_CGAP\_G4, NCI\_CGAP\_G06,  
NCI\_CGAP\_Brn25. These 21 libraries were pooled and a  
single-stranded DNA preparation of the resulting mixture  
was used as a tracer in a subtractive hybridization with  
a driver whose composition is detailed below:  
NCI\_CGAP\_Kid3 pool 1 LLAM 3334-3337, 3682-3683,  
3798-3803 (IMAGE Clonides 1322376-1323911, 1456008-1456775  
, 1500552-1502855); NCI\_CGAP\_Kid5 pool 1, LLAM 3338-3342  
, 3722-3725, 3776-3778 (IMAGE Clonides 1323912-1325831,  
1471368-1472903, 1492104-1493255); NCI\_CGAP\_Lu5 pool 1  
LLAM 3575-3582, 3851-3854 (IMAGE Clonides 1414920-1417991,  
1520904-1522439); NCI\_CGAP\_G4 pool 1 LLAM 3164-3167,  
3716-3720, 3733-3735 (IMAGE Clonides 1257096-1258631,  
1469064-1470983, 1475592-1476743); NCI\_CGAP\_Pr22 pool 1  
LLAM 2457-2459, 2758-2759, 3062-3068 (IMAGE Clonides  
985608-986759, 1101192-1101959, 1217928-1220615);

NCI\_CGAP\_Col0 pool 1 LLAM 2644-2653, 2871-2872 (IMAGE  
 Clonoids 1057416-1061255, 1144584-1145351). Subtraction  
 was performed as previously described (Bonaldi, Lennon &  
 Soares (1996)): Normalization and Subtraction: Two  
 Approaches To Facilitate Gene Discovery. Genome Research  
 6, 791-806.

TAG\_LIB=NCI\_CGAP\_GC4  
 TAG\_TISSUE=germ cell  
 TAG\_SEQ=AAATC\*

BASE COUNT 79 a 70 c 80 g 88 t 1 others  
 ORIGIN

Query Match 1.4%; Score 20; DB 9; Length 318;  
 Best Local Similarity 100.0%; Pred. No. 83;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1268 GCAGGCCGCGAGCTGGCCC 1287  
 Db 174 GCAGGCCGCGAGCTGGCCC 193

Search completed: August 14, 2003, 07:04:13  
 Job time : 2903 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2003, 01:55:34 ; Search time 422 Seconds  
(without alignments)  
9275.320 Million cell updates/sec

Title: US-09-701-229-1  
Perfect score: 1450  
Sequence: 1 cgtgctgacgctcgcca.....tgttagcgcgacgcgcac 1450

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 2552756 seqs, 1349719017 residues

Word size : 0  
Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :				N_Geneseq_19Jun03.*			
				1:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*		
				2:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*		
				3:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*		
				4:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*		
				5:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*		
				6:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*		
				7:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*		
				8:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*		
				9:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*		
				10:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*		
				11:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*		
				12:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*		
				13:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*		
				14:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*		
				15:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*		
				16:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*		
				17:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*		
				18:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*		
				19:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*		
				20:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*		
				21:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*		
				22:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*		
				23:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*		
				24:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*		
				25:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.*		

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	1450	100.0	1450	21	AZ47131 Pseudomonas aerugi
2	1347	92.9	1347	23	AAS54275 Pseudomonas aerugi
3	54	3.7	1200	23	AAS51550 Pseudomonas aerugi
4	36	2.5	116	23	AAS48736 Pseudomonas aerugi
5	36	2.5	1083	22	AAF86639 Pseudomonas aerugi
6	36	2.5	1083	23	AAS4276 Pseudomonas aerugi
7	24	1.7	1080	23	AAS89258 DNA encoding novel
8	24	1.7	1080	23	AAS89801 DNA encoding novel

9	24	1.7	1080	23	AAS90166 DNA encoding novel
10	24	1.7	1317	22	AAS81460 Escherichia coli p
11	24	1.7	1317	22	AAS52295 E. coli DNA for ce
12	24	1.7	2058	23	AAS73012 DNA encoding novel
13	24	1.7	3117	23	AAS69095 DNA encoding novel
14	24	1.7	3117	23	AAS89240 DNA encoding novel
15	24	1.7	4831	23	AAS94542 DNA encoding novel
16	24	1.7	4942	23	AAS89816 DNA encoding novel
17	24	1.7	4944	23	AAS89273 DNA encoding novel
18	24	1.7	4944	23	AAS90181 DNA encoding novel
c 19	22	1.5	27	22	AAF86641 Pseudomonas aerugi
c 20	20	1.4	507	21	AAF07985 Fusarium venenatum
c 21	20	1.4	3511	22	AAI59407 Human polynucleoti
c 22	20	1.4	3511	22	AAD09230 Human c-ski oncopr
c 23	20	1.4	42000	21	AAAG3349 Streptomyces globi
c 24	20	1.4	63164	21	AAAG3348 Streptomyces globi
c 25	20	1.4	109519	22	AAAG3348 Streptomyces globi
c 26	19	1.3	359	24	ABN23890 Microsporidia DNA
c 27	19	1.3	522	20	AAZ19400 Human ORFX polynuc
c 28	19	1.3	522	20	AAZ19400 M. tuberculosis re
c 29	19	1.3	966	23	AAS56152 Salmonella typhi D
c 30	19	1.3	1362	20	AAV59101 S. aureofaciens te
c 31	19	1.3	1857	24	ABQ90059 M. capsulatus gene
c 32	19	1.3	2169	22	AAI58060 Human polynucleoti
c 33	19	1.3	2932	13	AAQ25388 TXA2 receptor gene
c 34	19	1.3	2932	20	AAZ32161 Human thromboxane
c 35	19	1.3	2932	20	AAZ32162 Human endothelial
c 36	19	1.3	2932	23	AAS65879 DNA encoding novel
c 37	19	1.3	2932	24	ABK83654 Human cDNA differe
c 38	19	1.3	2932	25	ACAS6821 Human signalling p
c 39	19	1.3	2932	25	ABZ42829 Human thromboxane
c 40	19	1.3	3489	23	AAS54094 Pseudomonas aerugi
c 41	19	1.3	6390	22	AAF81370 Quorum sensing con
c 42	19	1.3	9333	22	AAK74093 Human immune/haema
c 43	19	1.3	14272	21	AAK75096 Nucleotide sequenc
c 44	19	1.3	21500	23	AAS99633 Propionibacterium
c 45	19	1.3	27541	22	AAD17185 Streptomyces nous

ALIGNMENTS

RESULT 1  
AAZ47131  
ID AAZ47131 standard; DNA; 1450 BP.  
XX AAZ47131;  
AC AAZ47131;  
DT 28-MAR-2000 (first entry)  
XX Pseudomonas aeruginosa murD gene.  
DE Pseudomonas aeruginosa murD gene.  
KW MurD; bacterial cell wall; biosynthesis; inhibitor; D-glutamate;  
KW UDP-N-acetylmuramyl-L-alanine precursor; ds.  
XX Pseudomonas aeruginosa.  
OS Pseudomonas aeruginosa.  
PN WO9961050-Al.  
XX  
PD 02-DEC-1999.  
XX  
PF 26-MAY-1999; 99WO-US11585.  
XX  
PR 29-MAY-1998; 98US-0087308.  
XX  
PA (MERI ) MERCK & CO INC.  
PI Pseudomonas aerugi  
PI Pseudomonas aerugi  
XX El-Sherbelini M, Azzollina B;  
DR WPT; 2000-072548/06.  
DR P-PSDB; AAY56047.  
XX  
PT New nucleic acid encoding the MurD protein of Pseudomonas aeruginosa,





23-MAY-2000; 2000US-206848P.  
26-MAY-2000; 2000US-207727P.  
23-OCT-2000; 2000US-242578P.  
27-NOV-2000; 2000US-253625P.  
22-DEC-2000; 2000US-257931P.  
16-FEB-2001; 2001US-269308P.  
(ELIT-) ELITRA PHARM INC.  
Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;  
Yamamoto RT, Xu HH;  
WPI; 2001-611495/70.  
P-PSDB; AAU36416.  
New polynucleotides for the identification and development of  
antibiotics, comprise sequences of antisense nucleic acids -  
Claim 27; Seq ID No 7912; 51pp; English.  
The invention relates to antisense inhibitors of genes essential to  
prokaryotic cellular proliferation, their use in identifying the  
genes, their use in the discovery of novel antibiotics, the essential  
genes themselves and the encoded proteins. The prokaryotes used are  
Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella  
pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The  
invention is also useful for the identification of potential new targets  
for antibiotic development. The antisense nucleic acids can also be used  
to identify proteins used in proliferation, to express these proteins,  
and to obtain antibodies capable of binding to the expressed proteins.  
The proteins can be used to screen compounds in rational drug discovery  
programmes. The antisense nucleic acid sequence is also useful to screen  
for homologous nucleic acids which are required for cell proliferation in  
a wide variety of organisms. The present sequence encodes an  
essential prokaryotic cellular proliferation protein.  
Note: The sequence data for this patent did not form part  
of the printed specification, but was obtained in electronic  
format directly from WIPO at  
ftp.wipo.int/pub/published\_pct\_sequences.  
Seq  
Sequence 1347 BP; 214 A; 453 C; 227 T; 0 other;

Query Match 92.9%; Score 1347; DB 23; Length 1347;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1347; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 51 ATGACCTGATCGCTCGACACACTTCGCGCATCGTTGCGCCTCGCGCAAGAGCGGCATG 110  
DB 1 ATGACCTGATCGCTCGACACACTTCGCGCATCGTTGCGCCTCGCGCAAGAGCGGCATG 60  
QY 111 TCCCTGTCGCGTACCTCGCGCGCGCGGCTTGCCTTTTCGCGTGTGCGATACCGAGAG 170  
DB 61 TCCCTGTCGCGTACCTCGCGCGCGCGGCTTGCCTTTTCGCGTGTGCGATACCGAGAG 120  
QY 171 AACCGCGCGAGCTGGCCACCTTGGTGCAGATATCCGAGTGGAGTGGTTCGCGC 230  
DB 121 AACCGCGCGAGCTGGCCACCTTGGTGCAGATATCCGAGTGGAGTGGTTCGCGC 180  
QY 231 GAACTCGACGCGAGTTCCTCTGCTCCCGCGCAACTCTATGTACGCGCGCGCTTTCG 290  
DB 181 GAACTCGACGCGAGTTCCTCTGCTCCCGCGCAACTCTATGTACGCGCGCGCTTTCG 240  
QY 291 CTGCGACCCCTTCGCTGTGTACAGCCCGCGCGAAGGCGTGCATCTCCGGTGACATC 350  
DB 241 CTGCGACCCCTTCGCTGTGTACAGCCCGCGCGAAGGCGTGCATCTCCGGTGACATC 300  
QY 351 GATCTCTTCGCGCGAGCGAAGCGCCCGGATCGTTCGCCATCACCCTGTTCCAAACCGAAG 410  
DB 301 GATCTCTTCGCGCGAGCGAAGCGCCCGGATCGTTCGCCATCACCCTGTTCCAAACCGAAG 360  
QY 411 AGCACCGTGACACCTTGTGGCGAATGGCGGTGGCGGAGCAAGCGTGTCCCGCTC 470  
DB 361 AGCACCGTGACACCTTGTGGCGAATGGCGGTGGCGGAGCAAGCGTGTCCCGCTC 420

QY 471 GCGGCAACCTCGGCACCCCGCGCTCGACCTGCTGCGCGAGACATCGAGCTGTACGTC 530  
DB 421 GCGGCAACCTCGGCACCCCGCGCTCGACCTGCTGCGCGAGACATCGAGCTGTACGTC 480  
QY 531 TTGGAGCTGTGAGCTTCCAGCTTGAACCTCGCATCGCTCAACCGCGAGGTGCGGACC 590  
DB 481 TTGGAGCTGTGAGCTTCCAGCTTGAACCTCGCATCGCTCAACCGCGAGGTGCGGACC 540  
QY 591 GTGCTGAACCTCAGGAAGACCATATGATGCTACGACGCGATGGCTGACTACACCTG 650  
DB 541 GTGCTGAACCTCAGGAAGACCATATGATGCTACGACGCGATGGCTGACTACACCTG 600  
QY 651 GCCAAGCACCGGATCTTCGCGGTCCCGCCAGGTCGTGGTGAATCGCGCCGATGCCCTG 710  
DB 601 GCCAAGCACCGGATCTTCGCGGTCCCGCCAGGTCGTGGTGAATCGCGCCGATGCCCTG 660  
QY 711 ACCGACCGCTGATGCGCGATACCGTGCCTGCTGCTGGTGGCTGAACAGCCGAC 770  
DB 661 ACCGACCGCTGATGCGCGATACCGTGCCTGCTGCTGGTGGCTGAACAGCCGAC 720  
QY 771 TTCAGGCTTTGCGCTGATCGAGGAAGACGCCAAGTGGCTGGCGTTCAGTTCGAC 830  
DB 721 TTCAGGCTTTGCGCTGATCGAGGAAGACGCCAAGTGGCTGGCGTTCAGTTCGAC 780  
QY 831 AAGCTGCTCCCGGTTGGCGAACTGAAGATCCGTGGCGCCCAACAATATTCCAACCGCTC 890  
DB 781 AAGCTGCTCCCGGTTGGCGAACTGAAGATCCGTGGCGCCCAACAATATTCCAACCGCTC 840  
QY 891 GCGCGCTGCGCTGGGCGCATGCGGTTCGCGCTGCGGTTCGAGCCCATCTGCGCGCGTG 950  
DB 841 GCGCGCTGCGCTGGGCGCATGCGGTTCGCGCTGCGGTTCGAGCCCATCTGCGCGCGTG 900  
QY 951 AAGCGGTTTCGCGCTGGCTCATCGCTGCCAGTGGGTACCGAGCGCGAGCGGCTGAGC 1010  
DB 901 AAGCGGTTTCGCGCTGGCTCATCGCTGCCAGTGGGTACCGAGCGCGAGCGGCTGAGC 960  
QY 1011 TACTACGACGATTTCAAGGCCACCAACGTCGCGCGCGCCCTGGCGCGCATCGAGGGGCTG 1070  
DB 961 TACTACGACGATTTCAAGGCCACCAACGTCGCGCGCGCCCTGGCGCGCATCGAGGGGCTG 1020  
QY 1071 GGTGCGGACATCGAGCGCAAGCTGTGCTGCTGCGCGCGGAGAGCGGAAGGGCGCGAT 1130  
DB 1021 GGTGCGGACATCGAGCGCAAGCTGTGCTGCTGCGCGCGGAGAGCGGAAGGGCGCGAT 1080  
QY 1131 TTCCATGACCTGCGCGAGCGGTGCGCGCTTCTCCCGGGCGGTGCTGCTGCGCGGT 1190  
DB 1081 TTCCATGACCTGCGCGAGCGGTGCGCGCTTCTCCCGGGCGGTGCTGCTGCGCGGT 1140  
QY 1191 GACGCGGGCTGATTTGCCAGGCACTGGGCAACGCGGTACCGCTGCTGCGGTTCGCAACG 1250  
DB 1141 GACGCGGGCTGATTTGCCAGGCACTGGGCAACGCGGTACCGCTGCTGCGGTTCGCAACG 1200  
QY 1251 CTGGACGAAGCAGTCCGCGAGCGCCCGAGCTGGCGCGGAGGCGATGCGGTGCTGTC 1310  
DB 1201 CTGGACGAAGCAGTCCGCGAGCGCCCGAGCTGGCGCGGAGGCGATGCGGTGCTGTC 1260  
QY 1311 TCGCGCGCTGCGCGAGCTGGACATGTTCAAGAACTTCGAAGAACGCGGAGCGCTGTC 1370  
DB 1261 TCGCGCGCTGCGCGAGCTGGACATGTTCAAGAACTTCGAAGAACGCGGAGCGCTGTC 1320  
QY 1371 GCCAAGCGGTAGAGGAGCTAGCGTGA 1397  
DB 1321 GCCAAGCGGTAGAGGAGCTAGCGTGA 1347  
RESULT 3  
AAS1550  
ID AAS1550 standard; DNA; 1200 BP.  
XX  
AC AAS1550;  
XX  
DT 13-FEB-2002 (first entry)

```

XX DE Pseudomonas aeruginosa DNA for cellular proliferation protein #135.
XX DE
XX DE
KW Antisense; ds; prokaryotic cellular proliferation gene;
KW Antibiotic; antibacterial; drug design.
XX OS
XX OS Pseudomonas aeruginosa.
XX PN WO200170955-A2.
XX PD 27-SEP-2001.
XX PF
XX PF 21-MAR-2001; 2001WO-US09180.
XX PR
XX PR 21-MAR-2000; 2000US-191078P.
XX PR 23-MAY-2000; 2000US-206848P.
XX PR 26-MAY-2000; 2000US-207727P.
XX PR 23-OCT-2000; 2000US-242578P.
XX PR 27-NOV-2000; 2000US-253625P.
XX PR 22-DEC-2000; 2000US-257931P.
XX PR 16-FEB-2001; 2001US-269308P.
XX PA (ELIT-) ELITRA PHARM INC.
XX PI
XX PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
XX PI Yamamoto RT, Xu HH;
XX XX
XX DR WPI; 2001-6111495/70.
XX XX
XX PT New polynucleotides for the identification and development of
XX PT antibiotics, comprise sequences of antisense nucleic acids -
XX PS
XX PS Claim 27; Seq ID No 4132; 511pp; English.
XX CC
XX CC The invention relates to antisense inhibitors of genes essential to
XX CC prokaryotic cellular proliferation, their use in identifying the
XX CC genes, their use in the discovery of novel antibiotics, the essential
XX CC genes themselves and the encoded proteins. The prokaryotes used are
XX CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
XX CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
XX CC invention is also useful for the identification of potential new targets
XX CC for antibiotic development. The antisense nucleic acids can also be used
XX CC to identify proteins used in proliferation, to express these proteins,
XX CC and to obtain antibodies capable of binding to the expressed proteins.
XX CC The proteins can be used to screen compounds in rational drug discovery
XX CC programmes. The antisense nucleic acid sequence is also useful to screen
XX CC for homologous nucleic acids which are required for cell proliferation in
XX CC a wide variety of organisms. The present sequence encodes an
XX CC essential prokaryotic cellular proliferation protein.
XX CC Note: The sequence data for this patent did not form part
XX CC of the printed specification, but was obtained in electronic
XX CC format directly from WIPO at
XX CC ftp.wipo.int/pub/published_pct_sequences.
XX SQ
XX SQ Sequence 1200 BP; 157 A; 384 C; 402 G; 257 T; 0 other;
XX
Query Match 3.7%; Score 54; DB 23; Length 1200;
Best Local Similarity 100.0%; Pred. No. 2e-14;
Matches 54; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1397 ATGCTGCTGCTGTTGGCCCTTCCGCTGCTGTTGAGCGGCGACGGCATC 1450
DB 1 ATGCTGCTGCTGTTGGCCCTTCCGCTGCTGTTGAGCGGCGACGGCATC 54
RESULT 4
AAS48736/C
ID AAS48736 standard; DNA; 116 BP.
XX
XX AAS48736;
XX
XX 13-FEB-2002 (first entry)
XX

```

```

XX DE Pseudomonas aeruginosa cellular proliferation inhibitory sequence #222.
XX DE
XX DE
KW Antisense; ss; prokaryotic cellular proliferation;
KW Antibiotic; antibacterial; drug design.
XX OS
XX OS Pseudomonas aeruginosa.
XX PN WO200170955-A2.
XX PD 27-SEP-2001.
XX PF
XX PF 21-MAR-2001; 2001WO-US09180.
XX PR
XX PR 21-MAR-2000; 2000US-191078P.
XX PR 23-MAY-2000; 2000US-206848P.
XX PR 26-MAY-2000; 2000US-207727P.
XX PR 23-OCT-2000; 2000US-242578P.
XX PR 27-NOV-2000; 2000US-253625P.
XX PR 22-DEC-2000; 2000US-257931P.
XX PR 16-FEB-2001; 2001US-269308P.
XX PA (ELIT-) ELITRA PHARM INC.
XX PI
XX PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
XX PI Yamamoto RT, Xu HH;
XX XX
XX DR WPI; 2001-6111495/70.
XX XX
XX PT New polynucleotides for the identification and development of
XX PT antibiotics, comprise sequences of antisense nucleic acids -
XX PS
XX PS Claim 1; Seq ID No 1313; 511pp; English.
XX CC
XX CC The invention relates to antisense inhibitors of genes essential to
XX CC prokaryotic cellular proliferation, their use in identifying the
XX CC genes, their use in the discovery of novel antibiotics, the essential
XX CC genes themselves and the encoded proteins. The prokaryotes used are
XX CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
XX CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
XX CC invention is also useful for the identification of potential new targets
XX CC for antibiotic development. The antisense nucleic acids can also be used
XX CC to identify proteins used in proliferation, to express these proteins,
XX CC and to obtain antibodies capable of binding to the expressed proteins.
XX CC The proteins can be used to screen compounds in rational drug discovery
XX CC programmes. The antisense nucleic acid sequence is also useful to screen
XX CC for homologous nucleic acids which are required for cell proliferation in
XX CC a wide variety of organisms. The present sequence is an antisense
XX CC oligonucleotide of the invention.
XX CC Note: The sequence data for this patent did not form part
XX CC of the printed specification, but was obtained in electronic
XX CC format directly from WIPO at
XX CC ftp.wipo.int/pub/published_pct_sequences.
XX SQ
XX SQ Sequence 116 BP; 27 A; 37 C; 44 G; 8 T; 0 other;
XX
Query Match 2.5%; Score 36; DB 23; Length 116;
Best Local Similarity 100.0%; Pred. No. 3.3e-06;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1415 CCCTTCCCGTCGCGCTGTTGAGCGGCGACGGCATC 1450
DB 116 CCCTTCCCGTCGCGCTGTTGAGCGGCGACGGCATC 81
RESULT 5
AAF86639
ID AAF86639 standard; DNA; 1083 BP.
XX
XX AAF86639;
XX
XX 17-JUL-2001 (first entry)
XX

```

DE Pseudomonas aeruginosa mray gene.  
XX  
KW Mray: phospho-N-acetylmuramoyl-pentapeptide translocase;  
KW peptidoglycan biosynthesis; bacterial cell wall; infection;  
KW drug screening; antibacterial; Pseudomonas; G+C rich bacterium; ds.  
XX  
OS Pseudomonas aeruginosa.  
XX  
FH Key Location/Qualifiers  
FT CDS 1..1083  
FT /\*tag= a  
FT /product= "Mray protein"  
FT /function= "Catalyses the transfer of N-acetylmuramic  
FT acid peptide to a bactoprenol phosphate  
FT carrier molecule in peptidoglycan  
FT biosynthesis"  
XX  
PN WO200125251-A1.  
XX  
XX 12-APR-2001.  
XX  
XX 29-SEP-2000; 2000WO-US27056.  
XX  
XX 04-OCT-1999; 99US-0157580.  
XX  
XX (MERI ) MERCK & CO INC.  
XX  
XX El-Sherbeini M, Azzolina B;  
XX  
XX WPI: 2001-308221/32.  
XX P-PSDB; AAB73487.  
XX  
XX New Mray gene and enzyme of Pseudomonas aeruginosa, useful in vitro  
PT assays for screening antibacterial compounds that target cell wall  
PT biosynthesis, particularly for screening antibiotics against  
PT Pseudomonas  
XX  
XX Claim 2; Fig 1; 22pp; English.  
XX  
XX This sequence represents the Pseudomonas aeruginosa mray gene, which  
CC encodes a phospho-N-acetylmuramoyl-pentapeptide translocase designated  
CC Mray. The Mray protein is involved in bacterial cell wall biosynthesis.  
CC It catalyses the first step of the membrane cycle of peptidoglycan  
CC biosynthesis, i.e., the transfer of an N-acetylmuramic acid peptide to a  
CC bactoprenol phosphate carrier molecule. The Mray protein is useful in  
CC in vitro assays to screen for antibacterial compounds that target cell  
CC wall biosynthesis. Inhibitors of the Mray protein are useful in  
CC preventing the growth of Pseudomonas and other G+C rich bacteria.  
CC Pseudomonas aeruginosa is an opportunistic pathogen which causes  
CC infections in patients with burns, neutropenia, or cystic fibrosis.  
CC Primers or probes derived from the mray gene are useful in nucleic acid  
CC amplification-based assays for detecting the presence of a polynucleotide  
CC encoding Pseudomonas aeruginosa Mray protein.  
XX  
XX Sequence 1083 BP; 164 A; 350 C; 333 G; 236 T; 0 other;  
SQ  
Query Match 2.5%; Score 36; DB 22; Length 1083;  
Best Local Similarity 100.0%; Pred. No. 2.6e-06;  
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2 GTGCTGATCGGCTCGCCACCTTGAAGTCGCTTGA 37  
|||||  
DB 1048 GTGCTGATCGGCTCGCCACCTTGAAGTCGCTTGA 1083  
RESULT 6  
AAS54276  
ID AAS54276 standard; DNA; 1083 BP.  
XX  
XX AAS54276;  
XX  
XX 13-FEB-2002 (first entry)  
XX  
XX

DE Pseudomonas aeruginosa DNA for cellular proliferation protein #407.  
XX  
KW Antisense; ds; prokaryotic cellular proliferation gene;  
KW antibiotic; antibacterial; drug design.  
XX  
OS Pseudomonas aeruginosa.  
XX  
PN WO200170955-A2.  
XX  
XX 27-SEP-2001.  
XX  
XX 21-MAR-2001; 2001WO-US09180.  
XX  
XX 21-MAR-2000; 2000US-191078P.  
XX 23-MAY-2000; 2000US-206848P.  
XX 26-MAY-2000; 2000US-20727P.  
XX 23-OCT-2000; 2000US-242578P.  
XX 27-NOV-2000; 2000US-253625P.  
XX 22-DEC-2000; 2000US-257931P.  
XX 16-FEB-2001; 2001US-269308P.  
XX  
XX (ELIT-) ELITRA PHARM INC.  
XX  
XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;  
XX Yamamoto RT, Xu HH;  
XX  
XX WPI: 2001-611495/70.  
XX P-PSDB; AAU36417.  
XX  
XX New polynucleotides for the identification and development of  
XX antibiotics, comprise sequences of antisense nucleic acids -  
XX  
XX Claim 27; Seq ID No 7913; 511pp; English.  
XX  
XX The invention relates to antisense inhibitors of genes essential to  
XX prokaryotic cellular proliferation, their use in identifying the  
XX genes, their use in the discovery of novel antibiotics, the essential  
XX genes themselves and the encoded proteins. The prokaryotes used are  
XX Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella  
XX pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The  
XX invention is also useful for the identification of potential new targets  
XX for antibiotic development. The antisense nucleic acids can also be used  
XX to identify proteins used in proliferation, to express these proteins,  
XX and to obtain antibodies capable of binding to the expressed proteins.  
XX The proteins can be used to screen compounds in rational drug discovery  
XX programmes. The antisense nucleic acid sequence is also useful to screen  
XX for homologous nucleic acids which are required for cell proliferation in  
XX a wide variety of organisms. The present sequence encodes an  
XX essential prokaryotic cellular proliferation protein.  
XX Note: The sequence data for this patent did not form part  
XX of the printed specification, but was obtained in electronic  
XX format directly from WIPO at  
XX ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 1083 BP; 163 A; 348 C; 336 G; 236 T; 0 other;  
SQ  
Query Match 2.5%; Score 36; DB 23; Length 1083;  
Best Local Similarity 100.0%; Pred. No. 2.6e-06;  
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2 GTGCTGATCGGCTCGCCACCTTGAAGTCGCTTGA 37  
|||||  
DB 1048 GTGCTGATCGGCTCGCCACCTTGAAGTCGCTTGA 1083  
RESULT 7  
AAS89258  
ID AAS89258 standard; cDNA; 1080 BP.  
XX  
XX AAS89258;  
XX  
XX 13-FEB-2002 (first entry)  
XX  
XX

DE DNA encoding novel human diagnostic protein #25062.  
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
 KW Homo sapiens.  
 XX WO200175067-A2.  
 XX 11-OCT-2001.  
 XX 30-MAR-2001; 2001WO-US08631.  
 XX 31-MAR-2000; 2000US-0540217.  
 XX 23-AUG-2000; 2000US-0649167.  
 XX (HYSE-) HYSEQ INC.  
 XX Drmanac RT, Liu C, Tang YT;  
 PI WPI; 2001-639362/73.  
 DR P-PSDB; ABG25071.  
 XX New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 PT Claim 1; SEQ ID No 25062; 103pp; English.  
 PS The invention relates to isolated polynucleotide (I) and  
 XX polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. AAS64197-AAS94564 represent novel human  
 CC diagnostic coding sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 XX Query Match 1.7%; Score 24; DB 23; Length 1080;  
 XX Best Local Similarity 100.0%; Pred. No. 0.67;  
 XX Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 537 CTGTCGAGCTTCAGCTGGAAACC 560  
 DB 40 CTGTCGAGCTTCAGCTGGAAACC 63  
 RESULT 8  
 AAS89801  
 ID AAS89801 standard; cDNA; 1080 BP.  
 XX  
 AC AAS89801;  
 XX 13-FEB-2002 (first entry)  
 XX  
 XX DNA encoding novel human diagnostic protein #25062.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
 XX Homo sapiens.  
 XX WO200175067-A2.  
 XX 11-OCT-2001.  
 XX 30-MAR-2001; 2001WO-US08631.  
 XX 31-MAR-2000; 2000US-0540217.  
 XX 23-AUG-2000; 2000US-0649167.  
 XX (HYSE-) HYSEQ INC.  
 XX Drmanac RT, Liu C, Tang YT;  
 PI WPI; 2001-639362/73.  
 DR P-PSDB; ABG25614.  
 XX New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 XX Claim 1; SEQ ID No 25605; 103pp; English.  
 XX The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. AAS64197-AAS94564 represent novel human  
 CC diagnostic coding sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 XX Query Match 1.7%; Score 24; DB 23; Length 1080;  
 XX Best Local Similarity 100.0%; Pred. No. 0.67;  
 XX Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 537 CTGTCGAGCTTCAGCTGGAAACC 560  
 DB 40 CTGTCGAGCTTCAGCTGGAAACC 63  
 RESULT 9  
 AAS90166  
 ID AAS90166 standard; cDNA; 1080 BP.  
 XX  
 AC AAS90166;  
 XX 13-FEB-2002 (first entry)  
 XX  
 XX DNA encoding novel human diagnostic protein #25970.  
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

```

XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US08631.
XX PR 31-MAR-2000; 2000US-0540217.
XX PR 23-AUG-2000; 2000US-0649167.
XX PA (HYSE-) HYSEQ INC.
XX PI Drmanac RT, Liu C, Tang YT;
XX DR WPI; 2001-639362/73.
XX DR P-PSDB; ABG25979.
XX PT New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensics, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity -
XX PS Claim 1; SEQ ID No 25970; 103pp; English.
XX CC The invention relates to isolated polynucleotide (I) and
XX CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX CC and gene mapping, and in recombinant production of (II). The
XX CC polynucleotides are also used in diagnostics as expressed sequence tags
XX CC for identifying expressed genes. (I) is useful in gene therapy techniques
XX CC to restore normal activity of (II) or to treat disease states involving
XX CC quantitating a polypeptide in tissue, as molecular weight markers and as
XX CC a food supplement. (II) and its binding partners are useful in medical
XX CC imaging of sites expressing (II). (I) and (II) are useful for treating
XX CC disorders involving aberrant protein expression or biological activity.
XX CC The polypeptide and polynucleotide sequences have applications in
XX CC diagnostics, forensics, gene mapping, identification of mutations
XX CC responsible for genetic disorders or other traits to assess biodiversity
XX CC and to produce other types of data and products dependent on DNA and
XX CC amino acid sequences. AAS64197-AAS94564 represent novel human
XX CC diagnostic coding sequences of the invention.
XX CC Note: The sequence data for this patent did not appear in the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 1080 BP; 267 A; 253 C; 299 G; 261 T; 0 other;

Query Match 1.7%; Score 24; DB 23; Length 1080;
Best Local Similarity 100.0%; Pred. No. 0.67;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 537 CTGTCGAGCTTCCAGCTGGAAACC 560
DB 40 CTGTCGAGCTTCCAGCTGGAAACC 63

RESULT 10
AAH81460
ID AAH81460 standard; DNA; 1317 BP.
XX AC AAH81460;
XX DT 21-SEP-2001 (first entry)
XX DE Escherichia coli protein encoding nucleotide sequence SEQ ID NO:259.
XX KW Escherichia coli; identification; proliferation; microorganism;
XX KW antimicrobial; antibacterial; antibiotic; gene therapy; diagnosis;
XX KW bacterial growth inhibition; ds.

```

```

OS Escherichia coli.
XX PN WO200148209-A2.
XX PD 05-JUL-2001.
XX PF 19-DEC-2000; 2000WO-US34419.
XX PR 23-DEC-1999; 99US-0173005.
XX PA (ELIT-) ELITRA PHARM INC.
XX PI Forsyth RA, Ohlsen KL, Zyskind JW;
XX DR WPI; 2001-457376/49.
XX DR P-PSDB; AAG98404.
XX PT Novel nucleic acids encoding proteins required for Escherichia coli
XX PT proliferation, useful for screening for antimicrobial agents -
XX PS Claim 9; Page 388-389; 596pp; English.
XX CC The present invention describes a purified or isolated nucleic acid
XX CC sequence (I) consisting essentially of one of the 93 nucleotide sequences
XX CC given in AAH81202 to AAH81294, where expression of the nucleic acid in a
XX CC microorganism is capable of inhibiting proliferation of a microorganism.
XX CC (I) have antibacterial and antibiotic activities, and can be used in
XX CC gene therapy. Expression of (I) in a microorganism inhibits proliferation
XX CC of the microorganism, and the manufactured antibiotic is useful for
XX CC reducing the activity or level of a gene product required for
XX CC proliferation of a microorganism in a subject, specifically humans. The
XX CC nucleic acids that inhibit bacterial growth or proliferation can be used
XX CC as antisense therapeutics for killing bacteria. In addition to
XX CC therapeutic applications, the nucleic acid sequences complementary to
XX CC sequences required for proliferation can be used as diagnostic tools.
XX CC For example, nucleic acid probes complementary to proliferation-required
XX CC sequences that are specific for particular species of microorganisms can
XX CC be used as probes to identify particular microorganism species in
XX CC clinical specimens. AAH81295 to AAH81487 encode the Escherichia coli
XX CC proteins given in AAG98239 to AAG98431, and AAH81488 to AAH81491
XX CC represent oligonucleotides, which are used in the exemplification of the
XX CC present invention.
XX SQ Sequence 1317 BP; 294 A; 323 C; 400 G; 300 T; 0 other;

Query Match 1.7%; Score 24; DB 22; Length 1317;
Best Local Similarity 100.0%; Pred. No. 0.66;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 537 CTGTCGAGCTTCCAGCTGGAAACC 560
DB 475 CTGTCGAGCTTCCAGCTGGAAACC 498

RESULT 11
AAS52295
ID AAS52295 standard; DNA; 1317 BP.
XX AC AAS52295;
XX DT 13-FEB-2002 (first entry)
XX DE E. coli DNA for cellular proliferation protein #17.
XX KW Antisense; ds; prokaryotic cellular proliferation gene;
XX KW antibiotic; antibacterial; drug design.
XX OS Escherichia coli.
XX PN WO200170955-A2.
XX PD 27-SEP-2001.

```

```

PF 21-MAR-2001; 2001WO-US09180.
XX
XX 21-MAR-2000; 2000US-191078P.
XX 21-MAR-2000; 2000US-206848P.
XX 23-MAY-2000; 2000US-207727P.
XX 23-OCT-2000; 2000US-242578P.
XX 27-NOV-2000; 2000US-253625P.
XX 22-DEC-2000; 2000US-257931P.
XX 16-FEB-2001; 2001US-269308P.
XX
XX (ELIT-) ELITRA PHARM INC.
XX
XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
XX Yamamoto RT, Xu HH;
XX WPI: 2001-611495/70.
XX P-PSDB; AAU34436.
XX
XX New polynucleotides for the identification and development of
XX antibiotics, comprise sequences of antisense nucleic acids -
XX
XX Claim 27; Seq ID No 5932; 511pp; English.
XX
XX The invention relates to antisense inhibitors of genes essential to
XX prokaryotic cellular proliferation, their use in identifying the
XX genes, their use in the discovery of novel antibiotics, the essential
XX genes themselves and the encoded proteins. The prokaryotes used are
XX Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
XX pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
XX invention is also useful for the identification of potential new targets
XX for antibiotic development. The antisense nucleic acids can also be used
XX to identify proteins used in proliferation, to express these proteins,
XX and to obtain antibodies capable of binding to the expressed proteins.
XX The proteins can be used to screen compounds in rational drug discovery
XX programmes. The antisense nucleic acid sequence is also useful to screen
XX for homologous nucleic acids which are required for cell proliferation in
XX a wide variety of organisms. The present sequence encodes an
XX essential prokaryotic cellular proliferation protein.
XX Note: The sequence data for this patent did not form part
XX of the printed specification, but was obtained in electronic
XX format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 1317 BP; 294 A; 323 C; 400 G; 300 T; 0 other;
XX
XX Query Match 1.7%; Score 24; DB 23; Length 1317;
XX Best Local Similarity 100.0%; Pred. No. 0.66;
XX Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 537 CTGTCGAGCTTCACGCTGGAAC 560
XX DB 475 CTGTCGAGCTTCACGCTGGAAC 498
XX
XX RESULT 12
XX AAS73012
XX ID AAS73012 standard; cDNA; 2058 BP.
XX
XX AC AAS73012;
XX
XX DT 13-FEB-2002 (first entry)
XX
XX DE DNA encoding novel human diagnostic protein #8816.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
XX OS Homo sapiens.
XX
XX PN WO200175067-A2.
XX
XX PD 11-OCT-2001.
XX
XX PF 30-MAR-2001; 2001WO-US08631.
XX
XX
XX 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
XX 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI: 2001-639362/73.
XX P-PSDB; ABG08825.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity -
XX
XX Claim 1; SEQ ID No 8816; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. AAS64197-AAS94564 represent novel human
XX diagnostic coding sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 2058 BP; 496 A; 526 C; 574 G; 462 T; 0 other;
XX
XX Query Match 1.7%; Score 24; DB 23; Length 2058;
XX Best Local Similarity 100.0%; Pred. No. 0.63;
XX Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 537 CTGTCGAGCTTCACGCTGGAAC 560
XX DB 1801 CTGTCGAGCTTCACGCTGGAAC 1824
XX
XX RESULT 13
XX AAS69095
XX ID AAS69095 standard; cDNA; 3117 BP.
XX
XX AC AAS69095;
XX
XX DT 13-FEB-2002 (first entry)
XX
XX DE DNA encoding novel human diagnostic protein #4899.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
XX OS Homo sapiens.
XX
XX PN WO200175067-A2.
XX
XX PD 11-OCT-2001.
XX
XX PF 30-MAR-2001; 2001WO-US08631.
XX
XX

```

PR	31-MAR-2000; 2000US-0540217.
PR	23-AUG-2000; 2000US-0649167.
XX	(HYSE-) HYSEQ INC.
XX	Dmanac RT, Liu C, Tang YT;
XX	PI
XX	DR WPI; 2001-639362/73.
XX	P-PSDB; ABG04908.
XX	New isolated polynucleotide and encoded polypeptides, useful in
XX	diagnostics, forensics, gene mapping, identification of mutations
PT	responsible for genetic disorders or other traits and to assess
PT	biodiversity -
PT	
XX	Claim 1; SEQ ID No 4899; 103pp; English.
XX	The invention relates to isolated polynucleotide (I) and
CC	polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC	polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC	and gene mapping, and in recombinant production of (II). The
CC	polynucleotides are also used in diagnostics as expressed sequence tags
CC	for identifying expressed genes. (I) is useful in gene therapy techniques
CC	to restore normal activity of (II) or to treat disease states involving
CC	(II). (II) is useful for generating antibodies against it, detecting or
CC	quantitating a polypeptide in tissue, as molecular weight markers and as
CC	a food supplement. (II) and its binding partners are useful in medical
CC	imaging of sites expressing (II). (I) and (II) are useful for treating
CC	disorders involving aberrant protein expression or biological activity.
CC	The polypeptide and polynucleotide sequences have applications in
CC	diagnostics, forensics, gene mapping, identification of mutations
CC	responsible for genetic disorders or other traits to assess biodiversity
CC	and to produce other types of data and products dependent on DNA and
CC	amino acid sequences. AAS64197-AAS94564 represent novel human
CC	diagnostic coding sequences of the invention.
CC	Note: The sequence data for this patent did not appear in the printed
CC	specification, but was obtained in electronic format directly from WIPO
CC	at ftp.wipo.int/pub/published_pct_sequences.
XX	
XX	Sequence 3117 BP; 749 A; 820 C; 856 G; 692 T; 0 other;
SQ	
Query Match	1.7%; Score 24; DB 23; Length 3117;
Best Local Similarity	100.0%; Pred. No. 0.6;
Matches	24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	537 CTGTGCAGCTTCACGCTGGAAACC 560
Db	
Db	373 CTGTGCAGCTTCACGCTGGAAACC 396
RESULT 14	
AAS89240	
ID	AAS89240 standard; cDNA; 3117 BP.
XX	AC
XX	AAAS89240;
XX	DT
XX	13-FEB-2002 (first entry)
DE	DNA encoding novel human diagnostic protein #25044.
XX	Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW	food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX	Homo sapiens.
OS	WO200175067-A2.
XX	PN
XX	11-OCT-2001.
XX	30-MAR-2001; 2001WO-US08631.
XX	31-MAR-2000; 2000US-0540217.
PR	23-AUG-2000; 2000US-0649167.
PR	(HYSE-) HYSEQ INC.
PA	

XX Drmanac RT, Liu C, Tang YT;  
 XX WPI: 2001-639362/73.  
 DR P-PSDB: ABG30355.  
 DR  
 XX New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity  
 XX  
 PS Claim 1: SEQ ID No 30346; 103pp; English.  
 XX  
 CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. AAS64197-RAS94564 represent novel human  
 CC diagnostic coding sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 4831 BP; 1180 A; 1249 C; 1353 G; 1049 T; 0 other;  
 Query Match 1.7%; Score 24; DB 23; Length 4831;  
 Best Local Similarity 100.0%; Pred. No. 0.58;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 537 CTGTCGAGCTTCAGCTGGAACCC 560  
 Db 4574 CTGTCGAGCTTCAGCTGGAACCC 4597  
 Search completed: August 14, 2003, 04:47:00  
 Job time : 425 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2003, 01:48:19 ; Search time 1500 Seconds  
(without alignments)  
2163.342 Million cell updates/sec

Title: US-09-701-229-1

Perfect score: 1450

Sequence: 1 cgtgctgctgcctgcgcga.....tgttgagcggcagcgcatc 1450

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1504479 seqs, 1118970152 residues

Total number of hits satisfying chosen parameters: 3008958

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Published\_Applications\_NA.:

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*
- 2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq.\*
- 3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq.\*
- 4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq.\*
- 5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq.\*
- 6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq.\*
- 7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq.\*
- 8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*
- 9: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq.\*
- 10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq.\*
- 11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq.\*
- 12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*
- 13: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq.\*
- 14: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq.\*
- 15: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq.\*
- 16: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq.\*
- 17: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1347	92.9	1347	9	US-09-815-242-7912
2	332.8	23.0	1317	9	US-09-815-242-9978
3	278.6	19.2	1317	9	US-09-741-669-259
4	278.6	19.2	1317	9	US-09-815-242-5932
5	180.6	12.5	9025608	14	US-10-156-761-1
6	180.2	12.4	1401	14	US-10-156-761-6097
7	173.8	12.0	1461	9	US-09-712-363-79
8	149.8	10.3	1314	9	US-09-815-242-7090
9	149.8	10.3	1830121	14	US-10-329-960-1
10	91.8	6.3	1352	10	US-09-974-300-1685
c 11	86	5.9	3309400	10	US-09-738-626-1
12	68.4	4.7	795	14	US-10-156-761-5987
c 13	68.4	4.7	9025608	14	US-10-156-761-1
c 14	67	4.6	640881	10	US-09-790-988-1
15	66.2	4.6	3759	14	US-10-156-761-2268
16	66	4.6	1959	14	US-10-156-761-2777

Sequence 1, Appli	10	US-09-815-242-7912	88421	4.6	66
Sequence 32, Appli	14	US-10-187-267A-32	786	4.5	17
Sequence 1, Appli	14	US-10-187-267A-1	36321	4.5	18
Sequence 3946, Ap	14	US-10-156-761-3946	1134	4.5	19
Sequence 4634, Ap	14	US-10-156-761-4634	1458	4.4	20
Sequence 6900, Ap	14	US-10-156-761-6900	1620	4.4	21
Sequence 2193, Ap	14	US-10-156-761-2193	1392	4.3	22
Sequence 2510, Ap	14	US-10-156-761-2510	2214	4.3	23
Sequence 7755, Ap	9	US-09-815-242-7755	1533	4.3	24
Sequence 21, Appli	11	US-09-793-708-21	5970	4.3	25
Sequence 7683, Ap	12	US-10-201-365-11	5970	4.3	26
Sequence 1880, Ap	9	US-09-815-242-7683	1197	4.2	27
Sequence 7, Appli	14	US-10-156-761-1880	1848	4.2	28
Sequence 7, Appli	10	US-09-861-289-7	1248	4.2	29
Sequence 7, Appli	10	US-09-860-846-7	1248	4.2	30
Sequence 7, Appli	11	US-09-988-384B-7	1248	4.2	31
Sequence 3, Appli	11	US-09-836-821-7	1248	4.2	32
Sequence 3, Appli	11	US-09-860-846-3	12441	4.2	33
Sequence 3, Appli	10	US-09-861-289-3	13613	4.2	34
Sequence 3, Appli	10	US-09-860-846-3	13613	4.2	35
Sequence 3, Appli	11	US-09-836-821-3	13613	4.2	36
Sequence 1, Appli	10	US-09-860-846-1	15872	4.1	37
Sequence 1, Appli	10	US-09-988-384B-1	15872	4.1	38
Sequence 1, Appli	11	US-09-836-821-1	15872	4.1	39
Sequence 6049, Ap	14	US-10-156-761-6049	1167	4.1	40
Sequence 5269, Ap	14	US-10-156-761-5269	966	4.1	41
Sequence 1710, Ap	14	US-10-156-761-1710	1092	4.1	42
Sequence 49, Appli	12	US-09-953-348-49	1545	4.1	43
					44
					45

#### ALIGNMENTS

#### RESULT 1

US-09-815-242-7912

; Sequence 7912, Application US/09815242

; Patent No. US20020061569A1

; GENERAL INFORMATION:

; APPLICANT: Haseibeck, Robert

; APPLICANT: Ohlsen, Karl L.

; APPLICANT: Zyskind, Judith W.

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John D.

; APPLICANT: Carr, Grant J.

; APPLICANT: Yamamoto, Robert T.

; APPLICANT: Xu, H. Howard

; TITLE OF INVENTION: Identification of Essential Genes in

; FILE REFERENCE: ELITRA.011A

; CURRENT APPLICATION NUMBER: US/09/815,242

; CURRENT FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; NUMBER OF SEQ ID NOS: 14110

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 7912

; LENGTH: 1347

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

; FEATURE:



Qy	1146	GAGCGGTGCGCGCCTTCTCCGCGCGCGGTGGTACTGCTTGGCCTGTAGAGCGCGGCTGATT	1205
Db	1072	CGCTATCTGACCGCGCATCGTATCCGCGCTGTATTGCTTTGGCGCGCATGGCGCGCAGCTT	1131
Qy	1206	GCCCAGGCACCTGGGCAACGGGGTACCCTGCTGGTGGCGCTCGCAACGCTGGAGCAAGCAGTC	1265
Db	1132	GCCGCATGC-----GTCCGGAATTCGCCCAACAGACTGAGACGATGGAAGAGCGCATG	1185
Qy	1266	CGGCGAGCCGCGCAGCTGGCCGCGGAAGCGCATCGGTGCTGTGTTGTTCGCCGCGCTGCGCG	1325
Db	1186	CGTTTGTGTGGCGCGCAGCTTCACCGCGGTGATATGTTGCTGTGTTTCCCGCGCTGCGCC	1245
Qy	1326	AGCCTGCACATCTTCAAGAACTTCGAAGAACCGGAGCCCTGTTTCGCCAAGACCGCTAGAG	1385
Db	1246	AGCCTGCATCAGTTTAAATAATTTTGACAACGGGGCGATGCTTTTACCCTCTGGCGAAG	1305
Qy	1386	GAGCTAGCTGCA 1397	
Db	1306	GAGTTAGTTGA 1317	
RESULT 3			
US-09-741-669-259			
; Sequence 259, Application US/09741669			
; Patent No. US2002022718A1			
; GENERAL INFORMATION:			
; APPLICANT: Forsyth, R. Allyn			
; APPLICANT: Ohlsen, Kari L.			
; APPLICANT: Zyskind, Judith W.			
; TITLE OF INVENTION: Genes identified as required for			
; TITLE OF INVENTION: proliferation of E. coli			
; FILE REFERENCE: ELIYPA.009A			
; CURRENT APPLICATION NUMBER: US/09/741.669			
; CURRENT FILING DATE: 2000-12-19			
; PRIOR APPLICATION NUMBER: US 60/173005			
; PRIOR FILING DATE: 1999-12-23			
; NUMBER OF SEQ ID NOS: 481			
; SOFTWARE: FastSeq for Windows Version 4.0			
; SEQ ID NO 259			
; LENGTH: 1317			
; TYPE: DNA			
; ORGANISM: Escherichia coli			
; FEATURE:			
; NAME/KEY: CDS			
; LOCATION: (1)...(1317)			
US-09-741-669-259			
Query Match 19.2%; Score 278.6; DB 9; Length 1317;			
Best Local Similarity 53.7%; Pred. No. 2.6e-56;			
Matches 709; Conservative 0; Mismatches 579; Indels 33; Gaps 5;			
Qy	80	CATCGTTGTGCGCCTCGGCAAGACGGGATGCTCCCTGGTGGGCTACCTGGGCGCGCGCGG	139
Db	27	CGTCATTATTCGCGCCTGGCGCCTCACCGGGCTTTCCTGGGTGGACTTTTCTCGCTCGCGG	86
Qy	140	CTTCGCTTTCGCGCTGCTGATACCCGAGAGAAACCCCGGAGCTGGCCACCCTCGCTGC	199
Db	87	TGTGACGCGCGCGCTTATFGATACGCTATGACACCGCTGGCTGGATAAATACCGG-	145
Qy	200	CCAGTATCCGCGAGGTGGAAGTGCCTTGGCGGGAACCTGCACGCCGAGTTCCTCTGCTCGC	259
Db	146	-----AAGCCGTAGAACCCACACGGGCAGTCTGAATGATGCTGATGGCGGC	197
Qy	260	CCGCGAACTCTATCTAGCCCCCGGCTTGTCTCGCTGGCACCCCTCGGCTGTGTAAGCGCGC	319
Db	198	AGATCTGATTGTCGCCAGTCCCGGTATTGCACTGGCGCATCCATCCTTAAGCGCTGCGCG	257
Qy	320	CGCGAAGGCGTGCATCTCCGCTGACATCGATCTCTCCCGCGGAGCGCAAGGCCCC	379
Db	258	TGATGCCGGAATCGAAATCTGTTGGCGATACGAGCTGTTCTGTCGGAAGCACAAGCAC	317
Qy	380	GATCGTCGCCATCACCGGTTTCCAAACGGCAAGACACCGTGACACACCTGGTGGCGGAAT	439

[illegible]

```

US-09-815-242-5932
; Sequence 5932, Application US/09815242
; Patent NO. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes.
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5932
; LENGTH: 1317
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1317)
US-09-815-242-5932

```

Query Match	Score	DB	Length
Best Local Similarity	19.2%	278.6	9
Matches	53.7%	Pred. NO. 2.6e-56;	
709; Conservative	0;	Mismatches	579;
		Indels	33;
		Gaps	5



```

Db 7361575 GGTGGCGCGCGACGTCGAGGAGGCTGCGCGCGCTCGCTTACGCTGGCACGCC 7361634
Qy 752 CGGCTGAACAGCGCGACTTCAAGGCTTTCGGGCTGATCGAGGAAGCGCCAGAAGTG 811
Db 7361635 CGGACCGTCCCACTCGGCGTCGTGGAGGCAATTCGTGCGACCGCCCTTCGTCGAG 7361694
Qy 812 GCTGGCGTTCCAGTTCGACAGCTGCTGCGGTTGGGAACTGAAGATCCGTGGCGCCA 871
Db 7361695 CCGCAGAGAAGCCCGAGGAGCTGGCGAGGTGGCGAGCTGATCCGCCCGCGCGCA 7361754
Qy 872 CAACATATTCAGCGCTGCGCGCGCTGGCGCTGGCGCTGAGCGCTGCGGCTGCGGTTGCA 931
Db 7361755 CAACATGCGCAACCGCTTGGCGAGCGCGCTGGCGCGCGCTTGGGCGCTGCGCGCTC 7361814
Qy 932 CGGCATGCTCGCGCGCTGAAGGCGTTTTCGGGCTGCTCATCGCTGCGCAAGTGGGTACG 991
Db 7361815 GCGGCTAGCGGAGGAGTGGCGGCTTCCGCGGAGCGCCACCGCATCGCGACAGTCGC 7361874
Qy 992 CGAGCGGAGGCGGTGAGCTACTACGAGCTATCAAGGCGCCACCAACGTCGCGCGCGCCCT 1051
Db 7361875 CGAGTGGAGCGGCTACCTACATCGAGAGCTCCAAAGCGCCACCAACGCGCGCGCA 7361934
Qy 1052 GCGGCGATCGAGGCGTGGTGCGGAC-----ATCGAGCGCAAGCTGGTGCTCGC 1105
Db 7361935 AGCTCTTTGGCGGCGTACGCGGTCCATCGTCTGGATCGCGCGCGGCTGCCCAAGGCGC 7361994
Qy 1106 CCGGCGAGAGCGCAAGCGCGCGGATTTCCATGACTCGCGGAGCGGTCGCGGCTTCTG 1165
Db 7361995 CTCCTTCGAGAGCTGTGGCGCAAGTGGCGACAGCGGCTCGGGGTGTGTGCTCATCG 7362054
Qy 1166 CCGGCGGCTGTACTGCTTGGCGCTGAGCGCGCGGCTGATTGCGC---AGGCACTGGGCAA 1222
Db 7362055 TGCGGATCGTGGCTGATCGGTGAAGCGCTTGGCGGACAGCGCGCGGAGTACCGCTCG 7362114
Qy 1223 CGCGTACCGCTGTGGCGCTGCGCAAGCTGGAGAGAGTCCGCGAGCGCGCGGAGCT 1282
Db 7362115 CGACTCGACCGGACGACACTGGGCGGATGCGCGGCTGTCGAGAGCGCGCGGCT 7362174
Qy 1283 GCGCGGAGGAGGATCGGCTGTGTGTCGCGCGCTGGCGGAGCTGGACATGTTCAA 1342
Db 7362175 CGCGTGGCGGAGACAGTACTGCTGCGCGCGCGCTGCTGCTGATGACATGTTGCG 7362234
Qy 1343 GAACCTCGAAGAGCGGCGGCTGTTGCGCAAGCGGTAGAGAGTACGCTGATGCTG 1402
Db 7362235 CAACATCAACAAGCGGCTGACGCGTTTCGCGGAGCGGTTTCGCGGACTCGGAGCTGATC 7362294
Qy 1403 TCGGTGTTGGCC 1415
Db 7362295 CCGCGCGCTCGCC 7362307

```

RESULT 6

```

US-10-156-761-6097
; Sequence 6097, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109

```

```

; SEQ ID NO 6097
; LENGTH: 1401
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1401)
US-10-156-761-6097

Query Match      12.48; Score 180.2; DB 14; Length 1401;
Best Local Similarity 49.68; Pred. No. 3.1e-33;
Matches 674; Conservative 0; Mismatches 628; Indels 57; Gaps 6;

Qy 80 CATGTTGTGGCGCTCGGCAAGAGCGCATGTCTCGTGGCTGCTACCTGGCGCGCGCGG 139
Db 27 CACGCTCGCGGGCTCGCGCTCTCCGGTATCCCGCGCGCGCGCTGCTGACAGCGCTCG 86
Qy 140 CTTGGCTTTTCGCGCTGTGATACCCGAGAGAACCCCGCGGAGTGCCCACTGCTGCTG 199
Db 87 CGCGGTGCTCACGGTCTCAACGACGCGGAGCGCTCCCGGCGCGCGGCGGGA 146
Qy 200 CCAGTATCCGACAGTGGAGTGGCTTTCGGGGAAGTTCGAGCGGCGGAGTTCCTCTCTCG 259
Db 147 TCTGGAGGCGCTCGGCATACCGTGCCTCGGTGACGGCGCGGACCTGCCCGAGGGAC 206
Qy 260 CCGCGAACTCTATGTACGCGCGCGCTTGTGCTGCGCACCCCTGCTGCTGTGCTACAG 319
Db 207 CGAGCTCATCTCATCAACACACCGCGCTGCGAGCGGACAAAGCGCTGTTCCGCGGCG 266
Qy 320 CCGGAAAGGCGTGGCGATCTCCGGTGATCATGCTCTTTCGCGCGCGAG-----GC 370
Db 267 CGAGGCGGGGCTCCCGGCTCTGGGCGGAGTTCGAATCTCGCTTGGCGGCTGCGCGCG 326
Qy 371 GAAGGCGCGGATCGTCCCATCACCGTTCACGCGAAGAGACACCGTGCACACCTTGGT 430
Db 327 ATCGCGCGCTGCTCGCGCTCAGGCGCACCAAGCGCAACACGACCGCTGCGAGTGTCT 385
Qy 431 GGGCGAAATGGCGGTGGCGCGGACAAAGCTGTGCGCTGGCGGCAACCTGCGGACCCCC 490
Db 387 CGCTCTCATCTTACAGCGCGCGGCGCTGCGACGCGCGGCTGCGCAACATCGCGCTCTC 446
Qy 491 GGGCGCTGACCTGCTGGCGCGGCA-----CATCGAGCTGTAGCTGTGGAGCTGTGAG 544
Db 447 GCTCTCTGAGCGGCTCTCGCGGAGGAGAGTACGAGCTCTCGCGGTGAACTCTCTCAG 506
Qy 545 CTTCCAGCTGGAACCTTCGATCGCTTCAACGCGGAGGTGGCGAGCTGCTGTAACGTCAG 604
Db 507 CTACGAGTGCAGTGGCGCGCTCCCTCGCGGCGGCGGCGGCGGCGGCTGCTGACATCGC 566
Qy 605 CGAAGACCATATGATCGCTACGACGCGATGGGTGATACCATCTGGCCCAAGCAACCGAT 664
Db 567 GCGCGACCATCTGACTGGCACGCGCTCCATGGAGGCGTACACCGCGCGGCGGCGGCTAT 626
Qy 665 CT-----TCCGCGGTGCGCGCGGCTGCTGTAAGTCCGCGCGGATGCGCT 709
Db 627 CTAGAGGGCAATCGGGTCCGCTCGCTTACAGCTGGCGGCGGCGGCGGCGGCGGAGCT 686
Qy 710 GACCGCGCGCTGATC-----GCGGATACCGTGGCGGCTGCTGCTGCTGCT 751
Db 687 GGTGGCGCGCGGAGCTCGAGGAGGCTGCGCGGCGGCTGCGGCTTACGCTGGCGACGCC 746
Qy 752 CCGCGCTGAACAAGCGGACTTCAAGGCTTTTCGGCGCTGATCGAGGAAGCGGCGGCAAGTG 811
Db 747 CGGACCGTCCCAACTCGGCGCTGCTGGAGGCGATTCTGCTGACCGCGCTTCTGCTGAGGA 806
Qy 812 GCTGGGCTTCCAGTTCGACAGCTGCTGCGCGTGGCGGAACTGAAGATCGTGGCGGCGCA 871
Db 807 CCGCGAAGAACGCGCGGAGGCTGGCGGAGGTGGCGGAGCTGATCGCTCCGCGCGCGCGCA 866
Qy 872 CAACATATTCAGCGCTCGCGGCTGGCGCTGGGCGCATCGGCTGCGGCTGCGGCTTCGA 931
Db 867 CAACATCGCAACGCGCTTCCGCGAGGCGGCTTGGCGGCGGCTTCCGCGGCTGCCCGCTC 926

```

QY 932 CGCCATGCTCGGCGCGCTGAGGCGTTTCCGGCCCTGGCTCATCGTCCAGTGGGTACG 991  
Db 927 GCGGTACGGGACGAGTCTCGGCGCTTCCGCGGACCGCCACCGCATCGCGACGTCG 986  
QY 992 CGAGCGGCGAGGCGTGTAGTACTACACGATTCCAAAGGCCACCAACGTCGGCGCCGCCCT 1051  
Db 987 CGACGTGGACGGGTCACTACATCAGCACTCCAAAGGCCACCAACGTCGGCGCGGA 1046  
QY 1052 GCGCGGATCGAGGGGCTGGTCCGAC-----ATCGAGGCGAAGCTGGTGTCTCGC 1105  
Db 1047 AGCCTTTTGGCGGCTACGGGTCCATCTGTGATCGCGCGGCTCGCCAAAGGCGC 1106  
QY 1106 GCGCGGAGCGCAAGGGCGCGATTTCATGACCTGCGGAGCGCGTGGTGTCTG 1165  
Db 1107 CTCCTTCGACGAGCTGGTGGCCAAAGTCGGCAGCGGCTCCGGGGTGTGTCTATCG 1166  
QY 1166 CCGGGCGGTGTACTCTTGGCGGTGACGCGCGGCTGATTGCC-----AGSCACTGGGCAA 1222  
Db 1167 TCGGGATCGTGGCTGATCCGTGAAGCCCTTGGCGACACGCGCCCGAGGTACCGTGT 1226  
QY 1223 CGCGGTACCGTGTGGCGGTGCGCAACGCTGGAAGAGAGTCCGCGAGCGCGCGAGCT 1282  
Db 1227 CGACCTCGACCGGACGACACTGGGGCGATGCCCGGCTGTCCAGGAGCGCGGCGCT 1286  
QY 1283 GCGCGCGAAGGAGTGGGTGCTGTTCGCGCGGCTGCGGAGCGCTGGACATGTTCAA 1342  
Db 1287 CGCGCTCGCGCGGACAGGTACTGCTGGCGCGGCTGTGCTCATGATGATGTTCG 1346  
QY 1343 GAACCTCGAAGACGCGGCGCTGTTCGCCAAAGCGGT 1381  
Db 1347 CAACACACACGCGGTGACGCGTTTCGGAGGCGGT 1385

## RESULT 7

US-09-712-363-79  
; Sequence 79, Application US/09712363  
; Patent No. US20020164588A1  
; GENERAL INFORMATION:  
; APPLICANT: Eisenberg, David  
; APPLICANT: Rotstein, Sergio H.  
; APPLICANT: Marcotte, Edward M.  
; TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND  
; FILE REFERENCE: 07419-032001  
; CURRENT APPLICATION NUMBER: US/09/712,363  
; CURRENT FILING DATE: 2000-11-13  
; PRIORITY APPLICATION NUMBER: PCT/US00/02246  
; PRIORITY FILING DATE: 2000-01-28  
; PRIORITY APPLICATION NUMBER: 60/179,531  
; PRIORITY FILING DATE: 2000-02-01  
; PRIORITY APPLICATION NUMBER: 60/117,844  
; PRIORITY FILING DATE: 1999-01-29  
; PRIORITY APPLICATION NUMBER: 60/118,206,  
; PRIORITY FILING DATE: 1999-02-01  
; PRIORITY APPLICATION NUMBER: 60/126,593  
; PRIORITY FILING DATE: 1999-03-26  
; PRIORITY APPLICATION NUMBER: 60/134,093  
; PRIORITY FILING DATE: 1999-05-14  
; PRIORITY APPLICATION NUMBER: 60/134,092  
; PRIORITY FILING DATE: 1999-05-14  
; PRIORITY APPLICATION NUMBER: 60/165,124  
; PRIORITY FILING DATE: 1999-11-12  
; PRIORITY APPLICATION NUMBER: 60/165,086  
; PRIORITY FILING DATE: 1999-11-12  
; NUMBER OF SEQ ID NOS: 292  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 79  
; LENGTH: 1461  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
US-09-712-363-79

Query Match

12.0%; Score 173.8; DB 10; Length 1461;

RESULT 8

Best Local Similarity 51.7%; Pred. No. 9.7e-32;  
Matches 499; Conservative 0; Mismatches 427; Indels 39; Gaps 3;  
QY 273 GTGAGCCCGCGCTGTGCTGCGCACCCCTGCTGCTGTCAGGCGCGCGGAAAGCGGTG 332  
Db 214 GCCAGTCCCGGCTTCTGCGCCGCAACCCGCTACTGCGCGCGCGCGGCGGCGGTG 273  
QY 333 CGCATCTCCGCTGACATCGATCTCTTCGCCCG-----CGAGGCG 371  
Db 274 CCGATCTGGGGTACGCTGGAGTTAGCTTGGCGCTAGACGACGCGGCTGCTACTACGACCG 333  
QY 372 AAGGCCCCGATCTGCGCCATCAGCGGTTCGAAGCGAAGAGACCGTGACACCCCTGCTG 431  
Db 334 CCGCGAGCTGGTGTGTGTCGCGCACCAAGCGCAAGAACCCACGACGCTGATGCTG 393  
QY 432 GCGAAATATGCGGTGCGCGGCAAGCTGTGCGCGTCCGCGCGCAACCTTCGCGCACCCCG 491  
Db 394 CACGCCATGCTGATCGCGGTGCGCGCGCGCTGCTGTGCGCAATATCGGACGTGCG 453  
QY 492 GCCTCGACCTGCTGCGCGGACGACATGAGCTGTAGCTGTGAGCTGTGAGCTTCCAG 551  
Db 454 GTGCTGGATGTGTCGACGAGCGCGGAGCTGTCGCCCTGGAGTTGTCCAGTTCCAG 513  
QY 552 CTGGAACCTTGCATCGCCTCAAGCGGAGGTGGCGACCGCTGTGACGTCAGCGTAGAC 611  
Db 514 CTGACTGGGCGCGCTGCTGCGCGCGGAGCGCGGCTGCTCAACATTCGCGAAGAC 573  
QY 612 CATATGATCGCTACGACGCGATGGCTGCTACCTGCGCGCAAGCACACCGATTCCTCCG 671  
Db 574 CACCTGGACTGGCATGCCAGATGGCGGATACACCGCGCAAGCGCGCGGTGCTGACC 633  
QY 672 GGTGCGCGCAGTGTGTGTAATGCGCGCGATGTCCTGACCGACCGCTGTGTCGCGAT 731  
Db 634 GCGGGGTAGCGGTGCGCGGCTGGATGACAGCGGAGCGCGCTGCTGAGACGGGTCA 693  
QY 732 ACCGTGCGGTGCTGCTGCG-----TTCGGCTGAAAGCGCGGCTTCAGGCTTTCGGCGT 788  
Db 694 CCGCGCAGGTGCGGGTTCGGCTTCGGCTGCGGCGAGCGCGCGCGGGAACCTGGCGGTG 753  
QY 789 ATCGAGGAGACGCGCGCAAGTGGCTGCGCTTCAGTTTCGACAACTGCTCCCGGTTTGGC 848  
Db 754 CCGGACCGCCACCTGGTTCGATCGCGCTTCTCCGACGACTTGACGCTGCTGCGGTGCG 813  
QY 849 GAACCTGAAGATCCGTGCGCGCCACAACTATTTCGAAGCGCTGCGCGCGCTGGCGCTGGC 908  
Db 814 TCGATACCGGTGCGAGTCCGGTTCGGCTGCTGAGCCCTGCGCGCGCGCGCTGGCC 873  
QY 909 CATGCGGTGCGCGCTGCGCTTCGAGCGCATGCTGCGCGCGCTGAGGCGTTTTCGGCGCTG 968  
Db 874 CGCTCGGTGCGGGTGCCTCGCGGTGCGATCGCGGACGCGGTGCTGCTTCGAGTGGCG 933  
QY 969 GCTCATCGTCCAGTGGGTACGCGAGCGCGAGGCGGTGAGCTACTACGACGATTCCCAAG 1028  
Db 934 CGACACCGCGCGGTTGGTGGCGCTTTCGCGACGCGATCACTACGTGAGACGACTCCAAG 993  
QY 1029 GCCACCAAGCTGCGCGCGCGCTGCGCGCGATCGAGGGGCTGGGTGCGGACATCGAGCGC 1088  
Db 994 GCCACCAAGCTGCGCGCGCGCTTTCGGTTCGTTG-----CATACCG 1038  
QY 1089 AAGCTGTGCTGCTCGCGCGGAGAGCGCAAGGCGCGCGATTTCCATGACCTGCGCGAG 1148  
Db 1039 AGGTGTGATGATCGCGGTGCGCTCAGGGCGCGCTGCTTACGCGCGAGTTGCG 1098  
QY 1149 CCGGTGCGCGCTTCTCGCGCGGCTGTGCTGCTGCGCGCTGACGCGCGGTGATTCG 1208  
Db 1099 GCGATGCGTTCGCGCGGTGCTGCGGTGCTGATCGCGCGGATCGCGAGCGGTGCG 1158  
QY 1209 CAGGC 1213  
Db 1159 GAGGC 1163



US-09-815-242-7090

; Sequence 7090, Application US/09815242  
; Patent No. US20020061569A1

; GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Kari L.

; APPLICANT: Zyskind, Judith W.

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John D.

; APPLICANT: Carr, Grant J.

; APPLICANT: Yamamoto, Robert T.

; APPLICANT: Xu, H. Howard

; TITLE OF INVENTION: Identification of Essential Genes in  
; Prokaryotes

; FILE REFERENCE: ELITRA.011A

; CURRENT APPLICATION NUMBER: US/09/815,242

; CURRENT FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; NUMBER OF SEQ ID NOS: 14110

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 7090

; LENGTH: 1314

; TYPE: DNA

; ORGANISM: Haemophilus influenzae

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (1)...(1314)

; US-09-815-242-7090

Query Match 10.3%; Score 149.8; DB 9; Length 1314;  
Best Local Similarity 51.0%; Pred. No. 4.1e-26;  
Matches 440; Conservative 0; Mismatches 407; Indels 16; Gaps 3;

QY 272 TGTGACGCGCGCTTGTGCGTCCGACCCCTGCGGTGTACAGCGCGCGGAAAGCGGT 331

DB 210 TATTAGCCAGGCTTGGGTAAACACACAGAAATTCACACCGCACTTAAAGCGGAGT 269

QY 332 CGCATCTCCGGTGACATCGATCTCTTCCCGCGGAGCGGAAGCGCCCGATCGTCCGCAT 391

DB 270 GGAAGTATCGCGCATATTGAATTATCTGCGCGCGACGACAAAGCAATTTGTGGGAT 329

QY 392 CACCGTTCCACCGAGACGCGGTACACACCTGCTGGCGGCAATGCGGTGGCGC 451

DB 330 TACAGTTCAATGGTAAAGTACCGTAACCTTATAGTTATGAAATGGCGAAAGCTGC 389

QY 452 GGACAAGCGTTCGCGCTCGCGGCAACCTCGGCACCCCGCGCTCGACCTGCTGGCGGA 511

DB 390 TGGTGTGAACTTGGTATGGCGGAAATATTTGGGATTCGCCCTTGTCTATGTTGAATGA 449

QY 512 CGACATCGAGTGTACGCTGTTGGAGCTGTGAGCTTCAGCTCGGAAACCTCGCATCGCCT 571

DB 450 AGATTGTGAACCTTTATGTACTAGAGCTTCTAGTTTTCAGCTTCGAGCAACTTATAGCTT 509

QY 572 CAACGCGAGTGGCGGACCGTGTGACCTGAGCGAGACACCATATGATCGCTACGAGG 631

DB 510 AAAAGCTGCGGACGAGCTGCTTGAACGCTGACTGAAGATCATATGATCGCTATATGGA 569

QY 632 CATGCGTACCTACCACTGGCGACGACCGGATCTTCGCGGTGCCCGCCAGGCTCGGT 691

DB 570 TTTAGAAGATTATCGCCCAAGCAAAATACCGCATTTATCATATGCTAAAGTAGTGTT 629

QY 692 GAATCGCGCGGATGCCCTGACCCGACCGCTGATCGCCGAT--ACCGTCCCTGCTGGTGG 749  
DB 630 GAACAATGAAGATAGGCTGACTTTTGGGAAACGAAATCAAGCGAAACATACCGTTTC 689  
QY 750 TTCGGCCTGAACAGCCCGGACTTCAAGGCTTTTCGGCCTGTATCGAGAGAGAGCGGCAGAG 809  
DB 690 TTTTGGGAAATAGTGGGATTTATTGGCTAAAAACT-----GAAAAATGCAAGCAA 741  
QY 810 TGGGTGGGTTCCAGTTCGACAAAGCTGCTGCGGTTGGCGACTCAAGATCCGTGCGCGC 869  
DB 742 TATTTAATGGTAAAGATGAAGTATTTTACCTTGTGAAGAGCTACATTTGTTGTCGTCG 801  
QY 870 CACAACATATCCACGCGCTCGCGCTGCGCGCTGCGGCTGCGGCTGCGGCTGCGGCTG 929  
DB 802 CATATATATGACACATTTTGGCAGCAACAGCATTTGGCAGCAAGCTATAGGTATTAATTA 861  
QY 930 GACGCGATGCTGGGCGCTGAAGCGGTTTTCGGCCTGCGCTGCTGCTGCGTGGGTTA 989  
DB 862 GATTCAATTCGTACCGCACTTCGTCAATTTCAAGGGTTAGATCATCTGTTTCAATTAGTG 921  
QY 990 CGCAGCGCGAGGCGTGAGCTACTACGACGATTCCAAAGGCCACCAACGTCGGCGCGCC 1049  
DB 922 CATCAAGCTAATGGCATTCGTTGGATTGACTCTAAGGCACAAATGTGGGGAGTACA 981  
QY 1050 CTGGCGCGGATCGAGGGGCTGGTGCCGACATCGACGCAAGCTGCTGCTGCTGCGCGC 1109  
DB 982 GTTGTGCTGCTGCGTGGGCT-----TTATATTAGGGTAAATTCGATTTGTTGCTAGGC 1035  
QY 1110 GGAGACGCGAAGGGCGCGCATTT 1132  
DB 1036 GGAGACGCGAAGGGCGCTGATTT 1058

RESULT 9

US-10-329-960-1

; Sequence 1, Application US/10329960

; Publication No. US20030099277A1

; GENERAL INFORMATION:

; APPLICANT: Fleischmann et al.

; TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus influenzae Rd Genome, F

; FILE OF INVENTION: Thereof, and Uses Thereof

; FILE REFERENCE: PB186P1

; CURRENT APPLICATION NUMBER: US/10/329,960

; CURRENT FILING DATE: 2003-01-02

; PRIOR APPLICATION NUMBER: US 09/643,990

; PRIOR FILING DATE: 2000-08-23

; PRIOR APPLICATION NUMBER: US 08/487,429

; PRIOR FILING DATE: 1995-06-07

; PRIOR APPLICATION NUMBER: US 08/426,787

; PRIOR FILING DATE: 1995-04-21

; NUMBER OF SEQ ID NOS: 1

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 1

; LENGTH: 1830121

; TYPE: DNA

; ORGANISM: Haemophilus influenzae

; FEATURE:

; NAME/KEY: misc.feature

; LOCATION: (4747)..(4747)

; OTHER INFORMATION: n equals a, t, g or c

; FEATURE:

; NAME/KEY: misc.feature

; LOCATION: (9921)..(9921)

; OTHER INFORMATION: n equals a, t, g or c

; FEATURE:

; NAME/KEY: misc.feature

; LOCATION: (10150)..(10150)

; OTHER INFORMATION: n equals a, t, g or c

; FEATURE:

; NAME/KEY: misc.feature

; LOCATION: (29298)..(29298)

; OTHER INFORMATION: n equals a, t, g or c



```
;/
;/ NAME/KEY: misc_feature
;/ LOCATION: (36543)..(36543)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/
;/ NAME/KEY: misc_feature
;/ LOCATION: (36551)..(36551)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/
;/ NAME/KEY: misc_feature
;/ LOCATION: (36636)..(36636)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/
;/ NAME/KEY: misc_feature
;/ LOCATION: (40808)..(40810)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/
;/ NAME/KEY: misc_feature
;/ LOCATION: (44416)..(44416)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/
;/ NAME/KEY: misc_feature
;/ LOCATION: (44905)..(44905)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/
;/ NAME/KEY: misc_feature
;/ LOCATION: (44975)..(44975)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/
;/ NAME/KEY: misc_feature
;/ LOCATION: (45593)..(45593)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/
;/ NAME/KEY: misc_feature
;/ LOCATION: (45732)..(45732)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/
;/ NAME/KEY: misc_feature
;/ LOCATION: (47036)..(47036)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/
;/ NAME/KEY: misc_feature
;/ LOCATION: (51334)..(51334)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/
;/ NAME/KEY: misc_feature
;/ LOCATION: (51602)..(51602)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/
;/ NAME/KEY: misc_feature
;/ LOCATION: (51786)..(51786)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/
;/ NAME/KEY: misc_feature
;/ LOCATION: (51805)..(51805)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/
;/ NAME/KEY: misc_feature
;/ LOCATION: (55369)..(55369)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/
;/ NAME/KEY: misc_feature
;/ LOCATION: (65309)..(65309)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/
;/ NAME/KEY: misc_feature
;/ LOCATION: (65313)..(65313)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/
;/ NAME/KEY: misc_feature
;/ LOCATION: (80024)..(80024)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
```

```
;/
;/ NAME/KEY: misc_feature
;/ LOCATION: (100091)..(100091)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/
;/ NAME/KEY: misc_feature
;/ LOCATION: (102696)..(102696)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/
;/ NAME/KEY: misc_feature
;/ LOCATION: (105121)..(105121)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/
;/ NAME/KEY: misc_feature
;/ LOCATION: (107248)..(107248)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/
;/ NAME/KEY: misc_feature
;/ LOCATION: (117136)..(117136)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/
;/ NAME/KEY: misc_feature
;/ LOCATION: (119750)..(119750)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/
;/ NAME/KEY: misc_feature
;/ LOCATION: (119924)..(119924)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/
;/ NAME/KEY: misc_feature
;/ LOCATION: (120038)..(120038)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/
;/ NAME/KEY: misc_feature
;/ LOCATION: (121344)..(121344)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/
;/ NAME/KEY: misc_feature
;/ LOCATION: (122167)..(122167)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/
;/ NAME/KEY: misc_feature
;/ LOCATION: (122336)..(122336)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/
;/ NAME/KEY: misc_feature
;/ LOCATION: (131340)..(131340)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/
;/ NAME/KEY: misc_feature
;/ LOCATION: (131360)..(131360)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/
;/ NAME/KEY: misc_feature
;/ LOCATION: (139910)..(139910)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/
;/ NAME/KEY: misc_feature
;/ LOCATION: (140398)..(140398)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/
;/ NAME/KEY: misc_feature
;/ LOCATION: (142750)..(142750)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/
;/ NAME/KEY: misc_feature
;/ LOCATION: (145058)..(145058)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/
;/ NAME/KEY: misc_feature
;/ LOCATION: (145171)..(145171)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/
;/ NAME/KEY: misc_feature
```

LOCATION: (145942)..(145942)  
 OTHER INFORMATION: n equals a, t, g or c  
 FEATURE:  
 NAME/KEY: misc\_feature  
 LOCATION: (147197)..(147197)  
 OTHER INFORMATION: n equals a, t, g or c  
 FEATURE:  
 NAME/KEY: misc\_feature  
 LOCATION: (150841)..(150841)  
 OTHER INFORMATION: n equals a, t, g or c  
 FEATURE:  
 NAME/KEY: misc\_feature  
 LOCATION: (152500)..(152500)  
 OTHER INFORMATION: n equals a, t, g or c  
 FEATURE:  
 NAME/KEY: misc\_feature  
 LOCATION: (152530)..(152530)

Query Match 10.3%; Score 149.8; DB 14; Length 1830121;  
 Best Local Similarity 51.0%; Pred. No. 2.9e-26;  
 Matches 440; Conservative 0; Mismatches 407; Indels 16; Gaps 3;  
 QY 272 TGTACGCGCGCTGTTCGCTGCGCACCCCTGCGTGTACAGCGCGCGCGAAGGCGT 331  
 DB 1203341 TATTAGCCAGCGCTTCGCGTAAACACCAAAATTCACACCGCACTTAAAGCGGAGT 1203400  
 QY 332 CGCATCTCCGTCGATCATCTCTTCGCCCGCGAGCGGAGCGCGCATCGTCGCCAT 391  
 DB 1203401 GGAAGTAATCGGATATTGAATATTCTGCGCGCAGCAGCAAGCCAAATGTGGGAT 1203460  
 QY 392 CACCGTTCCACGCGAAGACACCGCTGACACCCCTGTGGCGGAAATGCGGTGGCGC 451  
 DB 1203461 TACAGTTCAAATGTTAAAGTACCGTAACCTACTTATAGTTATGAATGGCGAAAGTCG 1203520  
 QY 452 GGACAAGCGTTCGCGCTGCGCGCAACCTCGCACCCCGCGCTCGACCTGCTGGCGCA 511  
 DB 1203521 TGGTGTGAAGTGTGTTGGCGGAAATATTGGGATTCGCCCTTGTCAATTTGAATGA 1203580  
 QY 512 CGACATCGAGCTGACGTGTGGAGCTTCGAGCTTCAGCTTGGAAACCTGCGATCGCT 571  
 DB 1203581 AGATTGTGAATTTATGTAAGCTTTCTAGTTTTCAGCTTGAGACAATTTAGCTT 1203640  
 QY 572 CAACGCCAGTGGGACCGTGTGAGCTGAGCGAAGACCAATATGATTCGCTACGAGG 631  
 DB 1203641 AAAAGCTCGCGCAGCGACTGCTTGAACGTGACTGAAGTATGATGATGCTATATGA 1203700  
 QY 632 CATGGCTGACTACCACTGGCCAGCACCGGATCTTCGCGGTGCGCCAGGCTGCTGGT 691  
 DB 1203701 TTTAGAAGATTATCGCCAAAGCAAAATTAACGCAATTTATCATTAAGTAGTGTGTT 1203760  
 QY 692 GAATCGCGCGATGCCCTGACCCGACCGCTGATCGCCGAT--ACCGTGGCTGCTGGTCG 749  
 DB 1203761 GAACAATGAAGTAGGCTGACTTTTGGGAAACGAAATCAAGGAAACATACCGTTTC 1203820  
 QY 750 TTCGCGCTGACACGCGGACTTCAAGCTTTTCGCGCTGATCGAGGAAAGCGCGCAAG 809  
 DB 1203821 TTTTCGCGAAATAGTCGGGATTATGGCTAAAACT-----GAAATGGCAAGCAA 1203872  
 QY 810 TGGCTGGGCTTCCAGTTCGACAAGCTGCTGCGGCTGCGCAACTGAAGTCCGTGGCGC 869  
 DB 1203873 TATTAATGGTAAAGATGAAGTATTTACCTGTGAAGAAGCTACATGGTGGTGGCG 1203932  
 QY 870 CACAACATTTCAACGCGCTCGCGCGCTGGCGCTGGGCCATCGGCTCGCGCTCCCGTTC 929  
 DB 1203933 CATAAATATGAACATTTTGGCAGCAACAGCATTTGGCACAAGCTATAGTATTAATTA 1203992  
 QY 930 GACGCCATGTCGCGCGCTGAAGCGGCTTTTCGCGCTGCTATCGCTGCGCAATGGGTA 989  
 DB 1203993 GATTCAATTCGTACCGCACTTCGTCAATTTCAAGGGGTAGATCATCTGTTTCAATTAGT 1204052  
 QY 990 CGCGAGCGGAGGCGTGAAGTACTACGATTCAGAGGCGCAACGCTGCGCGCGCC 1049  
 DB 1204053 CATCAAGCTAATGGCATTCGTTGGATTATGACTCTAAGCAACAAATATGGGGAGTACA 1204112

QY 1050 CTGGCGCGGATCGAGGGCTGGTCCGACATCGACGCAAGCTGTGCTCGCCGCGC 1109  
 DB 1204113 GTTGCTGCATTGGCTGGCT-----TTATATTAGGGTAAATTGCATTGTTGCTAGGC 1204166  
 QY 1110 GGAGACGGCAAGGGCGCGCATTT 1132  
 DB 1204167 GGAGACGGAAAGGGCGCTGATTT 1204189  
 RESULT 10  
 US-09-974-300-1685  
 ; Sequence 1685, Application US/09974300  
 ; Patent No. US20020146721A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Berka, Randy M.  
 ; APPLICANT: Clausen, Ib Groth  
 ; TITLE OF INVENTION: Methods For Monitoring Multiple Gene  
 ; TITLE OF INVENTION: Expression  
 ; FILE REFERENCE: 10085.500-US  
 ; CURRENT APPLICATION NUMBER: US/09/974,300  
 ; CURRENT FILING DATE: 2001-10-05  
 ; PRIOR APPLICATION NUMBER: 09/680,598  
 ; PRIOR FILING DATE: 2000-10-06  
 ; PRIOR APPLICATION NUMBER: 60/279,526  
 ; PRIOR FILING DATE: 2001-03-27  
 ; NUMBER OF SEQ ID NOS: 8481  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 1685  
 ; LENGTH: 1352  
 ; TYPE: DNA  
 ; ORGANISM: Bacillus licheniformis  
 US-09-974-300-1685

Query Match 6.3%; Score 91.8; DB 10; Length 1352;  
 Best Local Similarity 45.7%; Pred. No. 1.7e-12;  
 Matches 490; Conservative 0; Mismatches 562; Indels 21; Gaps 4;  
 QY 327 GCGCTGCGCATCTCCGGTGACATCGATCTTCGCCCGCGAGCGAAGCGCCGATCGTC 386  
 DB 289 GGTATCCGGTCTGGAGCGGTTGAACCTCGCTATCATCTGACAGATGCTCCGTTTCATC 348  
 QY 387 GCCATACCGGTTCCAAAGCGAAGACACCGTGACACCCCTGGTGGGGAATGCGGTG 446  
 DB 349 GGCATTACCGTTCAAAGCGAAGACGACACACGCTTGTATTGAAATGCTGAAG 408  
 QY 447 GCGCGGACAGCGTGTGCGCGTGGCGCAACCTCGCACCCCGCGCTCGACCTG--- 503  
 DB 409 GCCGATTGCGAAAGCGCTCGTTGCTGGGAATATFCGCTACGGCGGCAAGCGAGTGCC 468  
 QY 504 ---CTGGCGGACGACATCGAGCTGTACGTTGGAGCTGTGAGCTTCCAGCTGGAAC 560  
 DB 469 AATCAGCCAGCGGAGAGCAATGATGCTGACAGAGCTGTGAGTTTCAGCTCATGGG 528  
 QY 561 TGCATCGCTTCAACGCGGAGTGGCGACCGTGTGTAACGCTGACGAGCAAGACATATGAT 620  
 DB 529 ACTTATCAATTCAGACCGCAAGATTGGATTATTTAAACGTTATTGATGCCATCTCGAT 588  
 QY 621 CGCTACGACGCGATGGCTGACTACCACTTGCCCAAGCACCGGATCTCCCGGGTCCCGC 680  
 DB 589 TATCACCATTGCGCGCAAAATTTATGAGCTTCCCAACAGCAGGCTATTCGAAATCAATCG 648  
 QY 681 CAGTCTGTGGTGAATCGCGCGGATGCCCTGACCGGACCGCTGATCGCGGATACCGTGCG 740  
 DB 649 GAAACAGCGTTGCGGTTGCAATTTGGACGATGAGACAGCTGCTCCGCTCGCGAGTGT 708  
 QY 741 TGTGCTGTGGCTTGAACACCGGAGCTTCAAGGCTTTTCGCGCTGATGACGAGAGAC 800  
 DB 709 TCAAAAGCGGAGAGGTGTATTTTTCGCTCGCGCGACACTTGAAGCGGAGCGTGGCTG 768  
 QY 801 GCGCAGAGTGGCTGGGCTTCCAGTTCGACAAGCTGTGCGGTTGGGAACTGAAGATC 860  
 DB 769 AAGACGCGGCAATCATGTTTAAACGCGCAACCGGCTCATGCGCTGGAAGAGCGTCTGTTG 828



; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 5987  
; LENGTH: 795  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(795)  
US-10-156-761-5987

Query Match 4.7%; Score 68.4; DB 14; Length 795;  
Best Local Similarity 44.2%; Pred. No. 5.2e-07;  
Matches 282; Conservative 0; Mismatches 356; Indels 0; Gaps 0;  
QY 53 GAGCCTGATCGCCTCCGACCACTTCGGCATCGTTGTGGCCCTCGCAAGAGCGGCATGTC 112  
Db 156 GGGCCGGCGTACGGCGGACAAACGAGCGCCCTCGACGGCACCCCGCTCGTCGACGTGA 215  
QY 113 CTGTGTGCGTACCTGGCGCGCGGCTTGCCTTTCGCCGTGGTTCGATACCCGAGAGAA 172  
Db 216 CCGCGCGCAAGATCACTTGGCAGCGCCGGCCAGCTCTCGGGTATCCGATCCAGAA 275  
QY 173 CCGCGCGAGTGGCCACCCCTGCGTGCAGTATCCGAGTGGAAAGTGCCTTGGCGGA 232  
Db 276 ACTGCGCGCGGCTCGACGCTGCTGCCACCTCGCGCGCTCGAAGAGTTCATGATCCG 335  
QY 233 ACTGACGCGAGTTCCTCTGCTCGCGCGGAACTCTATCTACGCCCGCTTGTGCT 292  
Db 336 GCTGTGCGGAGTTCGGCTGGCCACGCGGCTGGAGGCGCGGGCGGCTGCGGT 395  
QY 293 GCGCACCCCTCGCTGTGTACAGCGCGCGGAAAGCTGGCGCATCTCCGGTGCATCGA 352  
Db 396 CCGCGGACCCGCTGGAGCAGCGCCCGAGCTCGCGGCTCTCCCTCGACTTCGACCC 455  
QY 353 TCTCTTCCCGCGGAGGAGCGCCGATGCTGCCATACCGGTTCACACGCGAAGAG 412  
Db 456 CCGCTGCGCGACGACGAGTTCGACCCCGCTCAACGGCGCCGAGTACGCCCGCTCAA 515  
QY 413 CACCGTACCACTCGTGTGGCGGAAATGGCGGTGGCGGACGACGCTGTCGCCGCG 472  
Db 516 CGCGCGCAGCGCGGAGGACCGCAAGATCTGCGCATGGGATCCGCTGCGCAAGGG 575  
QY 473 CGGCAACCTCGGACCGCGGCTGCGACTGCTGGCGGACGACATCGAGCTGTACGTGT 532  
Db 576 GGTCAACATGACGAGTTCGCGCTGAACGTGAATCCGAGACGCTCTCTTCGACAAAG 635  
QY 533 GGAGCTGCGAGCTTCCAGCTGGAACCTGCGATCGCTCAACCGCGAGGTGGCGACGT 592  
Db 636 CATCCGTCGGGATCCGCGACGCGGCGGTGACGTCTCCCTGACGAGCTGGGCGCGA 695  
QY 593 GCTGAACGTCAGCAAGACCAATATGGATCGCTAGCAGCGCATGCTGACTACACCTGGC 652  
Db 696 GGTCAACATCGCGAGGTCTCGCGGTGCGGAGCAAGCACTGCGGACGTAAGTGGAGAA 755  
QY 653 CAAGCAGCGATCTTCGCGGTGCCCGCCAGGTGCTGG 690  
Db 756 CGCGGACCTGAAGCGCGGAGGTGAACGCGGCTCGG 793

RESULT 13  
US-10-156-761-1/c  
; Sequence 1, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRO

; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156.761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 1  
; LENGTH: 9025608  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: misc.feature  
; LOCATION: (4187715)  
; OTHER INFORMATION: a, t, c, g, other or unknown  
US-10-156-761-1

Query Match 4.7%; Score 68.4; DB 14; Length 9025608;  
Best Local Similarity 44.2%; Pred. No. 3.3e-07;  
Matches 282; Conservative 0; Mismatches 356; Indels 0; Gaps 0;  
QY 53 GAGCCTGATCGCCTCCGACCACTTCGGCATCGTTGTGGCCCTCGCAAGAGCGGCATGTC 112  
Db 7237210 GGGCGGCGTACGGCGGACAAACGAGCGCCCTCGACGGCACCCCGCTCGACGTGA 7237151  
QY 113 CCGTGTGCGTACCTGGCGCGCGGCTTGCCTTTCGCCGTGGTTCGATACCCGAGAA 172  
Db 7237150 CCGCGGCGGCAAGATCACTTGGCAGCGCCCGGCGAGCTCTCGGGTATCCGATCCAGAA 7237091  
QY 173 CCGCGCGAGTGGCCACCTTGCCTGCCAGTATCCGAGTGGAAAGTGCCTTTCGCGGA 232  
Db 7237090 ACTCGCGCGCGTTCGAGCTGCTCGCCACCTCGCGCGCTCGAAGAGTTCATGATCG 7237031  
QY 233 ACTCGAGCGCGAGTTCCTCTGCTCGCGCGCGGAACTCTATGTCAGCCCGCGCTTGTGCT 292  
Db 7237030 GGTCTGCGCGAGTTCGCGCTGGCGCACCGCTGGAGGCGCGCGCTGCGGT 7236971  
QY 293 GCGCACCCCTCGCTGTGTACAGCGCGCGGAAAGGCTGGCATCTCCGGTGCATCGA 352  
Db 7236970 CCGCGGCGCGGCTGGAGCAGCGCGCGCTCGCGGCTCTCCCTCGACTTCGACCC 7236911  
QY 353 TCTCTTCCCGCGGAGGAGCGCCGATGCTGCCATCACCGTTCACACGCGAAGAG 412  
Db 7236910 CCGCTGCGCGACGAGTTCGACCCCGCTCAACGGCGCGGAGTACGCCCTCAA 7236851  
QY 413 CACCGTACCACTTCGCTGGCGGAAATGGCGGTGGCGGACGAGCTGTCGCCGCTCGG 472  
Db 7236850 CGCGGCGCGGCGGAGGACCGCAAGATCTGCGCATGGGATCGCGCTCGGCAAGGG 7236791  
QY 473 CGGCAACCTCGGCAACCCCGCTCGACCTGCTGGCGGACGACATCGAGCTGTACGTGT 532  
Db 7236790 CGTCACCATGCACGCTTTCGCCCTGAACGTGAATCCGAGACGCTCTCTTCGACAA 7236731  
QY 533 GGAGCTGTCGAGTTCAGCTGGAACCTTGGGATGCGCTCAACCGCGAGGTGGCGACCGT 592  
Db 7236730 CATCCGTCGGGATCGCGACGCGGCGGTGACGTCTCGCTGCGTACGAGTGGGCGCGA 7236671  
QY 593 GCTGAACGTCAGCAAGACCAATATGGATCGCTACGAGCGGATGCTGACTACACCTGGC 652  
Db 7236670 GGTCAACATCGCGAGGTCTTTCGCCGTGCGGAGCAAGCACTGCGGACGTAAGTGGAGAA 7236611  
QY 653 CAAGCAGCGATCTTCGCGGTGCCCGCCAGGTGCTGG 690  
Db 7236610 CGCGGACCTGAAGCGCGGAGGTGAACGCGGCTCGG 7236573

RESULT 14  
US-09-790-988-1/c  
; Sequence 1, Application US/09790988  
; Patent No. US20020127687A1  
; GENERAL INFORMATION:

```

; NAME/KEY: CDS
; LOCATION: (1)..(3759)
US-10-156-761-2268

Query Match      4.6%  Score 66.2:  DB 14;  Length 3759;
Best Local Similarity 46.6%  Pred. No. 1.6e-06;
Matches 212;  Conservative 0;  Mismatches 243;  Indels 0;  Gaps 0;

Qy      880  CCAACGGCGTCGCGCGCGCTGGCGCTGGGGCCATCGGTCGGCCCTGCCGTTTCGACGCCATGC 939
Db      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Qy      940  TCGCGCGGCTGAAGCGGTTTTCCGGCCCTGCTCATCGCTCCAGTGGGGTACGCGAGCGGC 999
Db      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Qy      545  CCACATTCGGCGCTGTGCGGGGGGTGCACGCCACACACCGCCGCGCGTTTCGGGGGAAGC 604
Db      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Qy      1000 AGGCGGTGAGCTACTACGACGATTCCAAAGCCACCAACGTCGGCGCGCGCCTGGCGGCGA 1059
Db      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Qy      605  TCGCGGAGGCTCGGGTGTCTGCTCTCAAGCGGCTCGCGAGCGCCACCCGCGACGGCG 664
Db      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Qy      1060 TCGAGGGGCTGGTGCGCGACATCGACGGCAAGCTGCTGCTCGCGCGCGGAGACGGCA 1119
Db      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Qy      665  ACACCGTGCTCGCCCTGGTAGCGGCGAGCGCCCTCGCGCAGGACGGCGACGGCGCGGCG 724
Db      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Qy      1120 AGGCGCGCGATTTCATGACTCTGGCGAGCGGTCGGCGCGCTTCTGCCGGGGCGGTGGTAC 1179
Db      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Qy      725  CGGCGCGCCCGACGGCGCGTCGAGGAGAAGGTCGTCGCGGCGCGCTCGCGCGCGGCC 784
Db      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Qy      1180 TGCTTGCCGTGAGCCGGGCTGATTCCCGAGGCACTGGCGCAACGCGGTACCGCTGGTGC 1239
Db      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Qy      785  GCCTCGGCCCGGAGGACATCAGTACGTCGAGGGCGACGGCACCGGCACCTCGCCGCGCG 844
Db      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Qy      1240 CGGTCGCAACGCTGGACGAAGCAAGTCGCGGAGCGCGCGAGCTGGCCCGCGAAGCGCATG 1299
Db      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Qy      845  GCTCCGTGAGTCCGGCGCGCTCGGCGCGCTCTTCGCGGAGTCGCACACCGAGGACGCTC 904
Db      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Qy      1300 CGGTGCTGTGTGTCGCGGCGCTGCGCGAGCCTGGAC 1334
Db      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Qy      905  CGGTGCTGTGTGCGGCTCGGTGCAGACCACTGGGC 939
Db      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

```

Search completed: August 14, 2003, 04:40:17  
Job time : 1536 secs

```

RESULT 13
US-10-156-761-2268
; Sequence 2268, Application US/10156761
; Publication NO. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 2268
; LENGTH: 3759
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:

```

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2003, 01:43:39 ; Search time 2899 Seconds  
(without alignments)  
12156.430 Million cell updates/sec

Title: US-09-701-229-1

Perfect score: 1450

Sequence: 1 cgtgctgctgcctgcga.....tgtgagcgcgcacggcatc 1450

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: em\_estba:\*

2: em\_esthum:\*

3: em\_estin:\*

4: em\_estmu:\*

5: em\_estov:\*

6: em\_estpl:\*

7: em\_estro:\*

8: em\_htc:\*

9: gb\_est1:\*

10: gb\_est2:\*

11: gb\_hic:\*

12: gb\_est3:\*

13: gb\_est4:\*

14: gb\_est5:\*

15: em\_estfun:\*

16: em\_estom:\*

17: em\_gss\_hum:\*

18: em\_gss\_inv:\*

19: em\_gss\_pin:\*

20: em\_gss\_vrt:\*

21: em\_gss\_fun:\*

22: em\_gss\_mam:\*

23: em\_gss\_mus:\*

24: em\_gss\_pro:\*

25: em\_gss\_rod:\*

26: em\_gss\_phg:\*

27: em\_gss\_vrl:\*

28: gb\_gsl1:\*

29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
c 1	643.6	44.4	892	29	BZ579116 msh2_6205
c 2	638.2	44.0	1445	29	BZ568959 pacs2-164
c 3	382.4	26.4	1171	29	BZ559806 pacs2-164
c 4	232.8	16.1	1194	29	BZ550283 pacs1-60_

c	5	217.2	15.0	928	29	BZ567557	pacs2-164
	6	179	12.3	452	10	BE184674	PM4-Ht068
	7	69.4	4.8	925	29	CNS0091P	AL053013 Drosophil
c	8	67.2	4.6	965	29	CNS070MR	AL423817 T3 end of
	9	66.4	4.6	935	29	CNS006XK	AL066051 Drosophil
c	10	63.6	4.4	1538	29	AG030607	AG030607 Pan trogl
c	11	62.6	4.3	925	29	CNS0091P	AL053013 Drosophil
	12	61.6	4.2	583	14	CD443549	CD443549 EL01N0427
	13	61.6	4.2	680	14	CD444231	CD444231 EL01N0437
	14	61.6	4.2	809	14	CD434773	CD434773 EL01N0328
	15	61	4.2	789	14	CD433407	CD433407 EL01N0308
	16	60.6	4.2	844	29	CNS0052P	AL056652 Drosophil
	17	60	4.1	602	14	CD435268	CD435268 EL01N0357
	18	60	4.1	629	14	CD445210	CD445210 EL01N0448
	19	60	4.1	638	14	CD435954	CD435954 EL01N0368
	20	60	4.1	639	14	CD442177	CD442177 EL01N0406
	21	60	4.1	646	14	CD445010	CD445010 EL01N0446
	22	60	4.1	654	14	CD434039	CD434039 EL01N0318
	23	60	4.1	661	14	CD435329	CD435329 EL01N0358
	24	60	4.1	678	14	CD444249	CD444249 EL01N0437
	25	60	4.1	680	14	CD445222	CD445222 EL01N0449
	26	60	4.1	683	14	CD442370	CD442370 EL01N0408
	27	60	4.1	691	14	CD443900	CD443900 EL01N0432
	28	60	4.1	704	14	CD434510	CD434510 EL01N0324
	29	60	4.1	720	14	CD444748	CD444748 EL01N0443
	30	60	4.1	758	14	CD440350	CD440350 EL01N0553
	31	60	4.1	771	14	CD440418	CD440418 EL01N0554
	32	60	4.1	772	14	CD434074	CD434074 EL01N0318
	33	60	4.1	778	14	CD434432	CD434432 EL01N0324
	34	60	4.1	791	14	CD438326	CD438326 EL01N0511
	35	60	4.1	797	14	CD437839	CD437839 EL01N0505
	36	60	4.1	799	14	CD434701	CD434701 EL01N0327
	37	60	4.1	808	14	CD438227	CD438227 EL01N0510
	38	60	4.1	809	14	CD433802	CD433802 EL01N0315
	39	60	4.1	810	14	CD438146	CD438146 EL01N0509
	40	60	4.1	820	14	CD437826	CD437826 EL01N0505
	41	60	4.1	825	14	CD434362	CD434362 EL01N0323
	42	60	4.1	828	14	CD440326	CD440326 EL01N0553
	43	60	4.1	846	14	CD439288	CD439288 EL01N0523
	44	60	4.1	854	14	CD437688	CD437688 EL01N0503
	45	60	4.1	859	14	CD439464	CD439464 EL01N0525

#### ALIGNMENTS

#### RESULT 1

#### BZ579116/c

#### LOCUS

#### DEFINITION

#### ACCESSION

#### VERSION

#### KEYWORDS

#### SOURCE

#### ORGANISM

#### REFERENCE

#### AUTHORS

#### TITLE

#### JOURNAL

#### COMMENT

BZ579116 msh2\_6205.y2 msh Pseudomonas aeruginosa genomic clone msh2\_6205, linear GSS 17-DEC-2002  
genomic survey sequence.

BZ579116 GI:27214177

GSS.

Pseudomonas aeruginosa

Pseudomonas aeruginosa

Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;

Pseudomonadaceae; Pseudomonas.

1 (bases 1 to 892)

Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,

Burns,J.L., Kaul,R. and Olsen,M.V.

Whole-Genome-Sequence variation among multiple isolates of

Pseudomonas aeruginosa library

J. Bacteriol., (2002) In press

Contact: Chris K. Raymond

Genome Center

University of Washington

Box 352145, Seattle, WA 98105-2145, USA

Tel: 2062216954

Fax: 2066857244

Email: craymond@u.washington.edu

Class: shotgun.

Location/Qualifiers

#### FEATURES

RESULT 2					
B2568959/c					
LOCUS	BZ568959	1445 bp	DNA	linear	GSS 17-DEC-2002
DEFINITION	pacs2-164_8171.y2 pacs2-164 pseudomonas aeruginosa genomic clone				



289	CTATTCCAACGCGCTCGCCGCGCTGGGCGCATGCGGTGGGCGTGCCTTCGACGC	230
935	CATGCTCGGCGCGCTGAAGCGTTTCCGGCCCTGGCTCATCGCTGCCAGTGGGTACGCGA	994
229	CATGCTCGGCGCGCTGAAGCGTTTCCGGCCCTGGCTCATCGCTGCCAGTGGGTACGCGA	170
995	GGCGCAGGCGTGAGCTACTACGACGATTCCAAGGCCACCAAGCTCGGCGCCGCGCTTGGC	1054
169	GGGCGAGGGCTGAGCTACTACGACGATTCCAAGGCCACCAAGCTCGGCGCCGCGCTTGGC	110
1055	GGCGATCGAGGGCGCTGGGTGCCGACATCGACGG	1087
109	GGCGATCGAGGGCGCTGGGTGCCGACATCGACGG	77

RESULT 3	BZ559806	1171 bp	linear	GSS 17-DEC-2002
LOCUS	BZ559806			
DEFINITION	pac32-164_1863.x1 pac32-164 Pseudomonas aeruginosa genomic clone			
ACCESSION	BZ559806			
VERSION	BZ559806.1			
KEYWORDS	GSS.			
SOURCE	Pseudomonas aeruginosa			
ORGANISM	Pseudomonas aeruginosa			
REFERENCE	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas.			
AUTHORS	1 (bases 1 to 1171)			
TITLE	Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M., Burns,J.L., Kaul,R. and Olsen,M.V.			
JOURNAL	Whole-Genome-Sequence variation among multiple isolates of Pseudomonas aeruginosa library			
COMMENT	J. Bacteriol., (2002) In press			
	Contact: Chris K. Raymond			
	Genome Center			
	University of Washington			
	Box 352145, Seattle, WA 98105-2145, USA			
	Tel: 2062216954			
	Fax: 2066857244			
	Email: craymond@u.washington.edu			
	Class: shotgun.			

```

class shotgun.
  FEATURES
    source
      Location/Qualifiers
        1..1171
          /organism="Pseudomonas aeruginosa"
          /mol_type="genomic DNA"
          /strain="2-164"
          /db_xref="taxon:287"
          /clone="pacs2-164_1863"
          /clone_lib="pacs2-164"
          /note="clinical isolate 2-164 Whole genomic shotgun
          library."
BASE COUNT      218 a   325 c   329 g   298 t   1 others
ORIGIN
Query Match      26.4%; Score 382.4; DB 29; Length 1171;
Best Local Similarity 93.9%; Pred. No. 4,9e-66;
Matches 398; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

```

875	CTATTCACAGCGCTCGCGCGCTGGCGCTGGGCCATGCGGTGGCGTGCCTTCACGC	934
QY		
47	CTATTCACAGCGCTCGCGAGCTGACACTGGSCCATGCGTCCGCTTTCGCTTTACGC	106
Db		
935	CATGCTCGGGCGCTCAAGCGCTTTTCGGCCTGGCTCATCGCTGCCAGTGGGTACGGA	994
QY		
107	CATGCTCGGGCGCTCAAGCGCTTTTCGCGACTGGCTCATTTGCTGCGACTGCGGA	166
Db		
995	GCGGCAAGGCGTGAGCTACTACACGATTTCCAAAGGCCACCAAGCTCGGCGCCGCTGGC	1054
QY		
167	GCGGCAAGGCGTGAGCTACTACACGATTTCCAAAGGCCACCAAGCTCGGCGCCGCTGGC	226
Db		
1055	GCGCATCGAGGGGCTGGTGCCGACATCGACGCCAAGCTGGTGCTGCTGCCCGCGGAGA	1114
QY		
227	GCGCATCGAGGGGCTGGTGCCGACATCGACGCCAAGCTGGTGCTGATCGCGCGGAGA	286
Db		

Qy	1115	CGGCAAGGGCCGCATTTTCATGACCTGCGCAGCCGGTTCGCGGGCGGT	1174
Db	287	CGGCAAGGGCCGCATTTTCATGACCTGCGCAGCCGGTTCGCGGGCGGT	346
Qy	1175	GGTACTGCTTTGGCCGTACGCCGGGCTGATTGCCAGGCACCTGGGCNAACGGGTACC	1234
Db	347	GGTACTGCTTTGGCGTGACACCTAGCTGATTGACCAAGGCACCTGGGCNAACGGGGACCG	406
Qy	1235	GGTGC GGCTCGCAACGCTGGACGAAGCAGTCCGGCAGCCGCGAGTGC GCCCGCAAGG	1294
Db	407	GGTGC GGCTCACAAACGCTGGACGAGCATTTCCGAGGCCGCGAGCTGCCCGCGAAG	466
Qy	1295	CGAT 1298	
Db	467	CGAT 470	

RESULT 4  
 BZ550283/c  
 LOCUS  
 DEFINITION  
 BZ550283  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Pseudomonas aeruginosa  
 Pseudomonas aeruginosa  
 Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
 Pseudomonadaceae; Pseudomonas.  
 1 (bases 1 to 1194)  
 Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,  
 Burns, J.L., Kaul, R. and Olsen, M.V.  
 Whole-Genome-Sequence Variation among Multiple Isolates of  
 Pseudomonas aeruginosa Library  
 J. Bacteriol., (2002) In press  
 Contact: Chris K. Raymond  
 Genome Center  
 University of Washington  
 Box 352145, Seattle, WA 98105-2145, USA  
 Tel: 2062216954  
 Fax: 2066857244  
 Email: craymond@u.washington.edu  
 Class: shotgun.

FEATURES	source	Location/Qualifiers
Class: shotgun:		
		1. .1194
		/organism="Pseudomonas aeruginosa"
		/mol_type="genomic DNA"
		/strain="1-60"
		/db_xref="taxon:287"
		/clone="pacsi-60_2584"
		/clone_lib="pacsi-60"
		/note="clinical isolate 1-60 whole genomic shotgun library."
BASE COUNT	233 a	349 c 298 g 314 t
ORIGIN		

	Query Match	16.1%	Score 232.8;	DB 29;	Length 1194;
	Best Local Similarity	83.0%;	Pred. No. 2.3e-36;		
	Matches 289;	Conservative 0;	Mismatches 57;	Indels 2;	Gaps 2;
Qy	1103	CGCCGGGAGACGGCAAGGCGCGCATATTCCATGACCTCGCGAGCCGTCGCGCGCTT	1162		
Dd	701	CCCCGGGAGAGAGACAACGGCCACCATTTCATGTCTTTGCCGACCCGTCGCGTGTT	642		
Qy	1163	CTCGCGGGCGGTGTACTGTCTTGCCCTGTACGCGGGCTGATTGCCCAGGCATCTGGGCAA	1222		
Dd	641	-TAACGAGCTGTGTCCTG-TTTGCCGGCCACCGATCTAAATCCCCAGCGCTGGGCAA	584		
Qy	1223	CGCGGTACCGTGTGTCGCTCGCAACGCTGGAGAACAGTAGTCGGGAGCGCCGCGAGCT	1382		
Dd	583	CACGGTCCATTGTGTGCGCGTCGACGCTTCGGAGCAAGCAGTAGTCGGGACGACGCTAGCT	524		

```

QY 1283 GGGCCCGAAGGCGATCGGTGCTGTTCGCGCGGCTCGCGAGCGCTGGACATGTTCAA 1342
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 523 GGGCCCGAAGGCGATCGGTGCTGTTCGCGCGGCTCGCGAGCGCTGGACATGTTCAA 1464
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1343 GAACTTCGAAGAACGGGACGCTGTCGCGAAAGCCGTAGAGGAGCTAGCGTGATGCTG 1402
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 463 GAACTTCGAAGAACGGGACGCTGTCGCGAAAGCCGTAGAGGAGCTAGCGTGATGCTG 404
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1403 TCGGTGTTTCGCGCCCTTCGCGCGCTGTTGAGCGCGGACGGCATC 1450
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 403 TCGGTGTTTCGCGCCCTTCGCGCGCTGTTGAGCGCGGACGGCATC 356
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 5
BZ567557/c 928 bp DNA linear GSS 17-DEC-2002
LOCUS
DEFINITION
pac2-164_7094.y2 pac2-164 Pseudomonas aeruginosa genomic clone
pac2-164_7094, genomic survey sequence.
BZ567557
VERSION
BZ567557.1 GI:27199005
KEYWORDS
SOURCE
Pseudomonas aeruginosa
ORGANISM
Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE
1 (bases 1 to 928)
Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
Burns,J.L., Kaul,R. and Olsen,M.V.
Whole-Genome-Sequence variation among multiple isolates of
Pseudomonas aeruginosa library
J. Bacteriol., (2002) in press
COMMENT
Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 20622216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.

FEATURES
    source
    1..928
    /organism="Pseudomonas aeruginosa"
    /mol_type="genomic DNA"
    /strain="2-164"
    /db_xref="taxon:287"
    /clone="pac2-164_7094"
    /clone_lib="pac2-164"
    /note="clinical isolate 2-164 whole genomic shotgun
    library."

BASE COUNT 238 a 228 c 236 g 226 t
ORIGIN
    Query Match 15.0%; Score 217.2; DB 29; Length 928;
    Best Local Similarity 71.4%; Pred. No. 2.8e-33;
    Matches 327; Conservative 0; Mismatches 128; Indels 3; Gaps 3;

QY 671 CGGTGCCCGCCAGGTCGTGGTGAA -TCGCGCGATGCCCTGACCCGACCGCTGATCGCGG 729
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 539 CGTTGCCCGTCAGGTCGTGGTTAAATTCGGCTGATGTCCTGACCCGCGCTGATCGTCC 480
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 730 ATACCGTCGCGTGGTGGTTCGCGCTGAACAAACGGGACTTCAAGGCTTTCGCGCTGA 789
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 479 ATCCCGTACCGTACGAGTGGCTGACGCTAGCAAGCGGTTTACAAAGCTTTCGACCTGG 420
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 790 TCGAGGAAGACGGCCAGAGTGGCTTCAGTTTCGACAAAGTGTGCGCGTTGGCG 849
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 419 TCCAAGAGTCTGACAGAGTGGCTGGGTTACGTTGACGAAATATATTCGGCTGCTG 360
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 850 AACTGAAGATCGTGGCGCCCAACAATATTCACACGCGCTCGCCGCTGGCGCTGGGCC 909
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 359 ATGTGAAGATCGTGGCGACGACGACTATTCACGCGCTTCCGCGCTGGCGTGGGCC 300
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 910 ATGCGGTGGCGCTGCCGTTCCGAC -GCCATGCTCGCGCGCTGAAGCGGCTTTCGCGCGCTG 968
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 299 ATGCGGTGCTGCGCGTCTTACAGCCATACATGGCGCACTCAAGGTGATCCACGTGT 240
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 969 GCTATCGCTGCGCAGTGGGTACGCGAGCGCGAGGCGTGAG-CTACTACGACCATTCACAA 1027
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 239 TCTCTATCGTGGCATGAGTACCTAACGACAGGGTCTGAGTATGTTCCGACGATTCAGCA 180
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1028 GGCCACCAACGTCGGCGCGCCCTGCGCGCATCGAGGGCGTGGTCCGACATCGACGCG 1087
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 179 GCCCGTCTTGGTGGTCACGCGCCCTGTTTCGATCGAGTGGATGGTGGTGGACATTTTCAT 120
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1088 CAACTGCTGCTGCTGCTGCGCGCGGAGACGCGAAGGCGG 1125
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 119 CAACCTCATGCTGCTGCGCAGACGACGACAAAGGGGG 82
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 6
BE184674
LOCUS
DEFINITION
PM4-HT0688-050500-002-d10 HT0688 Homo sapiens cDNA, mRNA sequence.
ACCESSION
BE184674
VERSION
BE184674.1 GI:8663858
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 452)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagal,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
JOURNAL
MEDLINE
20202663
PUBMED
10737800
COMMENT
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?cl=6t2=PM4-HT0688-050
500-002-d10&t3=2000-05-05&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 26
High quality sequence stop: 89.
Location/Qualifiers
    1..452
    /organism="Homo sapiens"
    /mol_type="mRNA"
    /db_xref="taxon:9606"
    /dev_stage="Adult"
    /clone_lib="HT0688"
    /note="Organ: head_neck; Vector: puc18; Site_1: SmaI;
    Site_2: SmaI; A mini-library was made by cloning products
    derived from ORESTES PCR (U.S. Letters Patent application
    No. 196,716 - Ludwig Institute for Cancer Research)
    profiles into the pUC 18 vector. Reverse transcription of
    tissue mRNA and cDNA amplification were performed under
    low stringency conditions."

BASE COUNT 91 a 140 c 133 g 88 t
ORIGIN
    Query Match 12.3%; Score 179; DB 10; Length 452;
    Best Local Similarity 73.3%; Pred. No. 9.4e-26;

```

BASE COUNT	120 a	61 c	61 g	172 t	511 others
ORIGIN	/note="end : "TET3"				
Query Match	4.8%; Score 69.4; DB 29; Length 925;				
Best Local Similarity	14.4%; Pred. No. 0.00058;				
Matches	58; Conservative	194; Mismatches	146; Indels	4; Gaps	1;
Qy	931	ACGCCATGCTCGCGCGCTCGAAGCGGTTTCCTGGCTGCTCATCGCTGCCATGGGTAC	990		
Db	524	ANNNNAGCSMCKGCKGTTTGGTTTTTSSGSGYKGCSSGGBSCSCSCSCSS 583			
Qy	991	GCAGCGGCAGGCGGTGAGCTACTACGACGATTCCAAAGGCACCAAGCTCGGCGCCGCC 1050			
Db	584	SCSCBCCCCSCSYCCSSBSBKCSFTSBSCCCSKSVCGFTSCSSSSCSSSSS 643			
Qy	1051	TGGCGGGATCGAGGGCTGGTGCCGACATCGACGGCAAGCTGTGTCTCGCCGGCG 1110			
Db	644	TSSSTSSSTSKSSSGSSSSSSSTTSKSTASGSGWSAGGSGTGSTSSSSSSS 703			
Qy	1111	GAGACGGCAAGGCGCGGATTCATGACCTCGCGAGCGCGTCTGTGCGGG 1170			
Db	704	TSYSSSVSGSKSTBSGSSBSSGSSSSSTWSBBSCSTSSSSSSSYSTSCCTC 763			
Qy	1171	CGGTGGTACGTGTGCGCTGACCGCGGCTGATGCCAGGCACTGGGCAACGCGGTAC 1230			
Db	764	CCSY----SYSSSTSSSSTSWGTSGSSSSSVGTSSSSDSTSTCCSCCYMGTCCSTYB 819			
Qy	1231	CGTGTGTGCGCTCGCAACCTGGACGAAGCACTCGCGGACGGCGCGAGCTGCCCGCG 1290			
Db	820	MBCTYTSFCGSSSSSKGVTCKCGCGCGSSSTNGMBGTSSACSSSSSSCSSSSVSSS 879			
Qy	1291	AAGCGATCGGTGCTGTTCGCGCGGCTCGCGAGCGTGG 1332			
Db	880	KSSASSSVSSSSGSGVSSNSASSKSSSGSVSSGSGSG 921			

QY	1291	AAGCGATGGTGCTTTGTCGGCGCCCTGCGGAGCCTGG	1332
DD	820	MBUITSICGSSSSSSGAGGVIKCGCGCGSSSSINGMEGTSACSSSSSSSCSSSSSSSSSS	873
Db	880	KSASSSSSSSSGSSGVSNNSSAKSSSSSGVSSSGSGSG	921

RESULT 8	CNS070MR	965 bp	linear	GSS 06-JUL-2001
LOCUS	CNS070MR/c			
DEFINITION	T3 end of clone AZ0AA010F07 of library AZ0AA from strain CBS 712 of <i>Kluyveromyces marxianus</i> , genomic survey sequence.			
ACCESSION	AL423817			
VERSION	AL423817.1			
KEYWORDS	GSS.			
SOURCE	GI:12207011			
ORGANISM	<i>Kluyveromyces marxianus</i>			
	<i>Kluyveromyces marxianus</i>			

ORGANISM  
Kluyveromyces fragilis  
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
Saccharomycetales; Saccharomycetaceae; Kluyveromyces.  
REFERENCE  
1 (bases 1 to 965)  
AUTHORS  
Souciet-J.L., Aigle, M., Artiguenave, F., Blandin, G.,  
Boloitin-Fukuhara, M., Bon, E., Brottier, P., Casaregola, S.,  
de-Montigny, J., Dujon, B., Durrens, P., Lepingle, A., Llorente, B.,  
Malpertuy, A., Neuvéglise, C., Ozler-Kalogeropoulos, O., Potier, S.,  
Saurin, W., Tekala, F., Toffano-Nioche, C., Wesolowski-Louvel, M.,  
Wincker, P. and Weissenbach, J.  
TITLE  
Genomic exploration of the hemiascomycetous yeasts: 1. A set of  
yeast species for molecular evolution studies  
JOURNAL  
FEBS Lett. 487 (1), 3-12 (2000)

JOURNAL MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

2 rue Gaston Cremieux, Cp 5706, 91057 EVRY cedex, FRANCE. (E-mail :  
segref@genoscope.cns.fr - Web : www.genoscope.cns.fr)  
This GSS is part of a random genomic sequencing program of thirteen  
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces  
exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,  
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces  
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia  
angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,  
Candida tropicalis and Yarrowia lipolytica. Genomic Inserts of 3 to  
5 kb were prepared and both extremities were sequenced. See  
keywords for description of this sequence and for the sequence of  
the other extremity of this insert.

COMMENT

FEATURES

source

1. .965

/organism="Kluyveromyces marxianus"

/mol\_type="genomic DNA"

/strain="CBS 712"

/variety="marxianus"

/db\_xref="taxon:4911"

/clone="AZ0AA010F0"

/clone\_lib="AZ0AA"

/note="end : 73"

complement(<355..>836)

/note="similar to CAB08672 [ murd ]

1 putative frameshift(s)"

/evidence=not\_experimental

BASE COUNT 199 a 275 c 310 g 159 t 22 others

ORIGIN

Query Match 4.6%; Score 67.2; DB 29; Length 965;

Best Local Similarity 47.6%; Pred. No. 0.0016;

Matches 226; Conservative 11; Mismatches 230; Indels 8; Gaps 2;

QY 916 TCGGCTCGCTGTCGACGCCATCTCGCGCGCTCAAGCGCTTTCCGCGCTGGCTCATC 975

DB 838 TCGGCGTGGCGCGCGCGCTGGTCTAGGGTCTCGCGGACTCCACCTTSGCGGATC 779

QY 976 GTGCGCATGGTACGCGAGCGGCGGCTGAGTACTACGAGATTCCAAAGGCCACA 1035

DB 778 GTATCGAGACCGTCCATAAAGCGGTGGCATCACSKGGKGCAGACTCCAAGGCCACW 719

QY 1036 ACGTCGCGCGCTGCGCGCATCGAGGGCTGGTGCCGAC-----ATCGACGCCA 1089

DB 718 ATYTTTTTTCGCTTCYCTCTGTCGCGGCTTCGAGACACATCTTGTGATCGCGGAG 659

QY 1090 AGCTGGTCTGCTCGCGCGGAGACGCGCGCGATTTCCATGACCTGCGCGAGC 1149

DB 658 GCCAGGCCAAGGCCACCTACTTTGAGACCTGGTCAACACCCAGCGGAGAGTGGCGG 599

QY 1150 CGGTGCGCGCTTTCGCGGCGGTG--GTACTGTTGGCGGTGACGCGCGGTGATTCG 1207

DB 598 AGCCATCGTGTGGGAACCGACCGTGTGCTGCTGCCAAGTCNTTGGCCGAGCACGCC 539

QY 1208 CCAGGCACTGGGCAACCGGTACCGCTGCTGGCGGTCCACGCTGGAGCAAGCATCCG 1267

DB 538 GCATGTCCTGCTGCTGATCATGATCTTTCGCGGAGATGCGGCAAGCGATTCG 479

QY 1268 CGAGCGCCGAGCTGGCGCGCGAGCGATCGGTGTGTCGCGCGCTCGCGGAG 1327

DB 478 CGAGGCTCCGAGATGGCGCGCGCGCGGACACCGTCTTGATGCTCGGGGTGGCGAG 419

QY 1328 CTGGACATGTTCAAGAACTTCGAAGACGCGGAGCGCTGTTGCGCAAGCCGTA 1382

DB 418 CTGGAATTTGGAATGWATATGATMMMGCGGTGACGATTTCTGTCAGCGCGCA 364

RESULT 9

CNS006XK

LOCUS

DEFINITION

Accession

CNS006XK 935 bp DNA linear GSS 03-JUN-1999  
Drosophila melanogaster genome survey sequence T7 end of BAC #  
BACR14N09 of RPCI-98 library from Drosophila melanogaster (fruit  
fly), genomic survey sequence.  
AL066051

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

AL066051.1 GI:4945019

GSS.

Drosophila melanogaster (fruit fly)

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephyridiidae; Drosophilidae; Drosophila.

1 (bases 1 to 935)

Genoscope.

Direct Submission

Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :

BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr

- Web : www.genoscope.cns.fr)

Determination of this BAC-end sequence was carried out as part of a

collaboration with the Berkeley Drosophila Genome Project (BDGP).

The BDGP is constructing a physical map of the Drosophila

melanogaster genome using these BACs. For further information

please see <http://www.fruitfly.org> The BDGP Drosophila

melanogaster BAC library was prepared by Kazutoyo Osoegawa and

Aaron Mammoser in Pieter de Jong's laboratory in the Department of

Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,

NY. The library is named RPCI-98 and was constructed by partial

EcoRI digestion of Drosophila DNA provided by the BDGP from the

isogenic strain y2; cn bw sp, the same strain used for the BDGP's

P1 and EST libraries. A more detailed description of the library

and how to order individual BAC clones, the entire library, or

filters for hybridization from the BACPAC Resource Center can be

found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

Location/Qualifiers

1. .935

/organism="Drosophila melanogaster"

/mol\_type="genomic DNA"

/db\_xref="taxon:7227"

/clone="BACR14N09"

/clone\_lib="RPCI-98"

/note="end : 17"

BASE COUNT 257 a 170 c 162 g 96 t 250 others

ORIGIN

Query Match 4.6%; Score 66.4; DB 29; Length 935;

Best Local Similarity 29.9%; Pred. No. 0.0023;

Matches 123; Conservative 108; Mismatches 180; Indels 1; Gaps 1;

QY 923 GCCGTTTCGACGCCATGTCGCGCGCTGAAGCGCTTTCCGCGCTGCTCATCGCTGCCA 982

DB 496 SSTGTMCDMMVCGCGCGSCMCRITSSGTCGCTCCMYMSSVSCCSCGTCGCGCSCTSK 555

QY 983 GTGGGTACGCGAGCGGAGCGGTGAGCTACTAGACGATTCCAAGGCCACCAACGTCGG 1042

DB 556 MCSCTYGCCKCGCGCTSCSSSSCCSBBSTCTCCTKCSGCTGSGCTGCCGGGG 615

QY 1043 CGCGCGCTGCGCGGATCGAGGGCTGGGTGCGGACATCGACGCAAGCTGTGCTGCT 1102

DB 616 SGCG 675

QY 1103 CGCGCGCGGAGACGGCAAGCGCGCGGATTTCCATGACCTGCGCGAGCGGTGCGCGCTT 1162

DB 676 SGCG 735

QY 1163 CTGCGCGCGGTGTGCTGCTGCGCGGTGAGCGCGCGGTGATTGCCAGGCACTGGGCA 1222

DB 736 CSCCG 795

QY 1223 CGCGCGCTACCGCTGTGCGCGCTCGCAAGCTGAGCAAGCTGCGCGAGCGCGCGCGCG 1282

DB 796 SGCG 854

QY 1283 GGCGCGCGAAGCGGATCGGTGCTGTTGTCGCCGCGCTGCGCGAGCTGGAC 1334

DB 855 SGSSCG 906

RESULT 10

[illegible]

Qy	884	CGCGCTCGCCGCGCTGCGCTGGGCCATGCGGTGCGCGTGCCTGCGCTTCAGACCCATGTCGG	943
Db	147	CGGGGCGTGTGAGAGAGTGGGCGCGCGCGTGCACACTGCTCGGTGGCGACCGCGT	206
Qy	944	CGCGCTGAAGCGCTTTTCGSGCCTGGCTCATCGCTGCCAGTGGGTACGGCAGCGCGAGGG	1003
Db	207	GGCGCTGGAGCCGGCGTCAAGTGTGTGGCGCTGCCGCCACTGCAAGGGCGGGCGGTACAA	266
Qy	1004	CGTGAGCTACTACGACGATTTCGAAGCGCCACCAAGCTCGGCGCGCGCTTGGCGCGATCGA	1063
Db	267	CGTGTGCGAGGACATGAAGTTCTTCGCCACCCCGCGGTGCACGGCTCGTGTGCGAACCA	336
Qy	1064	GGGCGTGGGTGCCGACATCGACCGCAAGCTGTGTGCTCGCGCGCGGAGACGGCAAGGS	1123
Db	327	GGTGGTGCACCCGGCGCACTGTGCTCAAGCTCCCGACGGGGTGAGCTGTGAGAGGS	386
Qy	1124	CGCCGATTTCCATGACTGCGCGAGCGGTGCGCGCTTCTGCGCGCGGTGGTACTGCT	1183
Db	387	CGCCATGTGCGAGCCGCTGAGCGTGGCGCTGCAACGCGTGCCGCGCGCGGGGTGGGCC	446
Qy	1184	TGGCGGTGAGCGCGGCGTGATTCGCCAGGCACTTGGGCAACCGGTTACCGGTGTGTGCGCGT	1243
Db	447	CGAGACGGGGTGTCTGTGTGTGGCGCGCGCCCATCGGCTGGTGTGCTGTGTGCGCGC	506
Qy	1244	CGCAACGCTGGAGAAAGCACTCGCGGACGCGCGGAGCTGGCCCGCGAAGGCGATCGGT	1303
Db	507	GCGGGCTTCGGCGCGCGCGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	566
Qy	1304	GCTGTTGTGCGCGCC	1319
Db	567	GGCCAGGTGCTGGGC	582

[illegible]

CD444231	GI:31359874	
CD444231.1		
ACCESSION	EST.	
VERSION	KEYWORDS	
KEYWORDS	SOURCE	
ORGANISM		
	Zea mays	
	Zea mays	
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.	
REFERENCE	1 (bases 1 to 680)	
AUTHORS	Lai,J., Dey,N., Kim,C.S., Becraft,P., Larkins,B., Linton,E. and Messing,J.	
TITLE	Sequencing of the maize endosperm ESTs	
JOURNAL	Unpublished	
COMMENT	Contact: Lai, Jinsheng Dr. Joachim Messing's lab Waksman Institute, Rutgers University 190 Frelinghuysen Rd., Piscataway, NJ 08854, USA Tel: 732-445-3801 Fax: 732-445-5735 Email: jlai@waksman.rutgers.edu Seq primer: T3.	

FEATURES	seq primer: 13.	Location/Qualifiers
source	1..680	
	/organism="Zea mays"	
	/mol_type="mRNA"	
	/cultivar="W22"	
	/db_xref="taxon:4577"	
	/tissue_type="Endosperm of 7-23DAP"	
	/clone_lib="Endosperm_4"	
	/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI"	
BASE COUNT	85 a 207 c 288 g	99 t 1 others
ORIGIN		
Query Match	4.2%;	Score 61.6; DB 14; Length 680;
Best Local Similarity	46.3%;	Pred. NO. 0.019;



Query Match	4.2%	Score 61;	DB 14;	Length 789;
Best Local Similarity	46.6%	Pred. No. 0.026;		
Matches 196;	Conservative 0;	Mismatches 225;	Indels 0;	Gaps 0;

  

QY	899	GGCGCTGGGCATGCTGGCTCGCGCTTCGACGCATGCTCGGGCGCCTGAAGCGCTT	958
Db	23	GGAGGTGGGGCGCGGTGACCTTCTGTCGTGGCGACCGCGTGGCGTGGAGCCGGG	82
QY	959	TTCCGGCGCTGGCTCATGCTGTGCAGTGGGTACCGAGCGGCAGGGCGTGAAGTACTACGA	1018
Db	83	CGTCACTGCTGGCGCTGCGCCACTGCAAGGCGGGCGGTACAACCTGTGCGAGGACAT	142
QY	1019	CGATTCCAAAGGCCAACCAAGCTCGCGCCCGCCCTGGCGGCGATCGAGGGCGTGGTGCCGA	1078
Db	143	GAAGTTCTTCGCACACCCCGCGGTGCACGGCTCGCTGGCGACCAAGTGGTGACCCGGC	202
QY	1079	CATTCGACGGCAAGCTGGTGCTGCTCCCGCGGAGACGGCAAGGGCGCGGATTTCATGA	1138
Db	203	CGACCTGTGCTTCAAGCTCCCGACGGGTGAGCCTGGAGAGGGCGCCATGTGCGAGCC	262
QY	1139	CTTGCGCGAGCGGTGCGCGCTTTCGCGCGGGGTGACTCTGTGGCGGTGACGCCGG	1198
Db	263	GCTGACCGTGGCGCTCCACGCGTGCGCCCGCGGGGGTGGGCCCGCAGACGGCGCTGTCT	322
QY	1199	GCTGATTGCCAGGCACTGGGCAACCGGTACCGCTGGTGCAGCTGCGCAACGCTGGACGA	1258
Db	323	CGTGGTGGCGCGCGGCCCATCGGCCCTGGTGTGCTGTGGCGCGCGGGCCCTTTCGGCGC	382
QY	1259	AGCAGTCCGCGAGCGCGCGAGCTGCCCCCGGAGGCGATGCGGTGCTGTGTGCGCGCGC	1318
Db	383	GGCGCGCTGTGGTCTGGATGTGGACGACCACCGCCCTGGCGCGTGGCCAGGTGCGTGGG	442
QY	1319	C 1319	
Db	443	C 443	

Search completed: August 14, 2003, 04:12:24  
Job time : 2903 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2003, 00:43:39 ; Search time 422 Seconds  
(without alignments)  
9275.320 Million cell updates/sec

Title: US-09-701-229-1

Perfect score: 1450

Sequence: 1 cgtgctgctgctgctgcca.....tggtgagcggcagcgcatc 1450

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_19Jun03.\*  
1: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*  
25: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	1450	100.0	1450	21	AAZ47131
2	1347	92.9	1347	23	AAS34275
3	332.8	23.0	1317	23	AAS56341
4	278.6	19.2	1317	22	AAH81460
5	278.6	19.2	1317	23	AAS52295
6	180	12.4	1380	21	AAZ53175
7	180	12.4	69936	21	AAH81479
8	180	12.4	349980	21	AAF21607
					Pseudomonas aerugi
					Pseudomonas aerugi
					Salmonella typhi D
					Escherichia coli p
					E. coli DNA for'ce
					Neisseria meningit
					N. meningitidis pa
					Neisseria meningit

9	180	12.4	1437668	21	AAH81490	N. meningitidis B
10	173.8	12.0	1461	22	AAH52025	Mycobacterium tube
c 11	173.8	12.0	4411529	22	AAJ99682	Mycobacterium tube
c 12	172.2	11.9	4403785	22	AAJ99683	Mycobacterium tube
13	165.6	11.4	1335	25	ABZ41528	N. gonorrhoeae nuc
14	158	10.9	4942	23	AAH89816	DNA encoding novel
15	158	10.9	4944	23	AAS89273	DNA encoding novel
16	158	10.9	4944	23	AAS90181	DNA encoding novel
17	157.2	10.8	1080	23	AAS89258	DNA encoding novel
18	157.2	10.8	1080	23	AAS89801	DNA encoding novel
19	157.2	10.8	1080	23	AAS90166	DNA encoding novel
20	152	10.5	2058	23	AAS73012	DNA encoding novel
21	152	10.5	4831	23	AAS94542	DNA encoding novel
22	151.4	10.4	3117	23	AAS69095	DNA encoding novel
23	151.4	10.4	3117	23	AAS89240	DNA encoding novel
24	149.8	10.3	1314	23	AAS53453	Haemophilus influe
25	149.8	10.3	1830121	17	AAJ42083	Haemophilus influe
c 26	119	8.2	96109	22	AAF28548	Genomic fragment #
27	110	7.6	2703	23	AAS64579	DNA encoding novel
28	110	7.6	2703	23	AAS66661	DNA encoding novel
29	91.8	6.3	1352	24	ABK74394	Bacillus lichenifo
c 30	86	5.9	34980	22	AAH68531	C glutamicum codin
c 31	79.8	5.5	356	21	AAZ53173	Neisseria meningit
c 32	78.4	5.4	356	21	AAZ53174	Neisseria meningit
33	76	5.2	9515	15	AAQ55145	Pseudomonas aerugi
34	76	5.2	9515	25	ABZ77356	Nucleotide sequenc
35	76	5.2	9515	25	AAJ51844	Pseudomonas aerugi
c 36	74.6	5.1	37856	21	AAJ11992	S. cellulosum DNA
c 37	73.6	5.1	356	21	AAZ53172	Neisseria gonorrhe
38	73.6	5.1	44377	18	AAJ78508	Platenolide syntha
39	73.6	5.1	44377	18	AAJ80414	Platenolide syntha
c 40	73.2	5.0	450	21	AAZ53171	Neisseria meningit
41	68.8	4.7	8169	19	AAJ26609	Actinomadura hibis
42	67.4	4.6	1119	24	ABQ90030	M. capsulatus gene
c 43	67	4.6	640681	24	ABA92787	Buchnera sp. genom
c 44	66.8	4.6	450	21	AAZ53169	Neisseria gonorrhe
45	66.8	4.6	53789	19	AAV21187	Anycolatopsis medi

#### ALIGNMENTS

##### RESULT 1

AAZ47131  
ID AAZ47131 standard; DNA; 1450 BP.

XX AAZ47131;

AC AAZ47131;

XX 28-MAR-2000 (first entry)

DE Pseudomonas aeruginosa murd gene.

DE Murd; bacterial cell wall; biosynthesis; inhibitor; D-glutamate;

KW UDP-N-acetylmuramyl-L-alanine precursor; ds.

XX Pseudomonas aeruginosa.

XX WO9961050-A1.

XX 02-DEC-1999.

XX 26-MAY-1999; 99WO-US11585.

XX 29-MAY-1998; 98US-0087308.

XX (MERI ) MERCK & CO INC.

XX El-Sherbeini M, Azzolina B;

XX WPI: 2000-072548/06.

XX P-FSDB; AAY56047.

PT New nucleic acid encoding the MurD protein of Pseudomonas aeruginosa,

PT used to identify specific inhibitors -

PS Claim 4; Page 8-9; 35pp; English.

XX This sequence represents the MurD gene from *Pseudomonas aeruginosa*.  
 CC MurD is an enzyme involved in biosynthesis of bacterial cell walls,  
 CC so inhibiting it inhibits bacterial growth (by preventing  
 CC MurD-catalyzed addition of D-glutamate to an alanine residue in the  
 CC UDP-N-acetylmuramyl-L-alanine precursor). Cells that express murD  
 CC protein can be used: (a) to identify inhibitors of MurD, and (b) for  
 CC recombinant production of MurD, which is used to raise specific  
 CC antibodies (potential therapeutic inhibitors). Fragments of the murD  
 CC gene can be used as probes and primers to detect MurD-encoding sequences  
 CC (to identify polymorphic murD genes or to detect *P. aeruginosa*), also  
 CC as antisense modulators. The murD inhibitors are useful for treating or  
 CC preventing bacterial infections, including those caused by *P. aeruginosa*.

XX Sequence 1450 BP; 227 A; 487 C; 487 G; 249 T; 0 other;

Query Match 100.0%; Score 1450; DB 21; Length 1450;

Best Local Similarity 100.0%; Pred. No. 3.7e-252;

Matches 1450; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGTGTGATCGGCTCGCCACCTTGAAGCTGCTTGGAGCAGAGAGCATGAGCCTGA 60  
 DB 1 CGTGTGATCGGCTCGCCACCTTGAAGCTGCTTGGAGCAGAGAGCATGAGCCTGA 60  
 QY 61 TCCGCTCCGACCACTTCCGACATGTTGCGGCTCGGCAAGAGCGCATGCTCCGTGTC 120  
 DB 61 TCCGCTCCGACCACTTCCGACATGTTGCGGCTCGGCAAGAGCGCATGCTCCGTGTC 120  
 QY 121 GCTACCTGGCGCGCGGCTTCCCTTTCGCGCTGCTGATACCCGAGAAACCCGCCGG 180  
 DB 121 GCTACCTGGCGCGCGGCTTCCCTTTCGCGCTGCTGATACCCGAGAAACCCGCCGG 180  
 QY 181 AGCTGGCCACCTTCGCTGCCAGTATCCGAGGTGGAAGTGCCTTGGGGAACCTCGACG 240  
 DB 181 AGCTGGCCACCTTCGCTGCCAGTATCCGAGGTGGAAGTGCCTTGGGGAACCTCGACG 240  
 QY 241 CCGAGTTCCTCTCTCCGCGCGGAACTCTATGTCAGCCCGGCTTGTGCTCGGCACCC 300  
 DB 241 CCGAGTTCCTCTCTCCGCGCGGAACTCTATGTCAGCCCGGCTTGTGCTCGGCACCC 300  
 QY 301 CTCGGCTGTACAGCGCGCGGAAAGGGGTGCGCATCTCCCGTGATCATGATCTCTTCG 360  
 DB 301 CTCGGCTGTACAGCGCGCGGAAAGGGGTGCGCATCTCCCGTGATCATGATCTCTTCG 360  
 QY 361 CCGCGAGGCGAAGCGCCGATCGCCATCACCGGTTCCACGCGAAGACCGTGA 420  
 DB 361 CCGCGAGGCGAAGCGCCGATCGCCATCACCGGTTCCACGCGAAGACCGTGA 420  
 QY 421 CCACCTGTGTGGCGAAATGGCGGTGGCGCGGACAAAGGTGTGCGCGCGGCAACC 480  
 DB 421 CCACCTGTGTGGCGAAATGGCGGTGGCGCGGACAAAGGTGTGCGCGCGGCAACC 480  
 QY 481 TCGGCAACCGCGCGCTCGACCTGCTGGCGACACATCATGAGTGTACGTGTTGGAGCTGT 540  
 DB 481 TCGGCAACCGCGCGCTCGACCTGCTGGCGACACATCATGAGTGTACGTGTTGGAGCTGT 540  
 QY 541 CGAGCTTCAGCTGGAACCTCGGATCGGCTCAACCGCGGAGTGGCGACCTGCTGAACG 600  
 DB 541 CGAGCTTCAGCTGGAACCTCGGATCGGCTCAACCGCGGAGTGGCGACCTGCTGAACG 600  
 QY 601 TCAGCAAGACCATATGGATCGGTACGAGCGCATGCTGACTACCACTGGCGCAACGACC 660  
 DB 601 TCAGCAAGACCATATGGATCGGTACGAGCGCATGCTGACTACCACTGGCGCAACGACC 660  
 QY 661 GGATCTTCGCGGTGCCCGCAGGTGCTGTGAATCGCGCCGATGCCCTGACCCGACCGC 720  
 DB 661 GGATCTTCGCGGTGCCCGCAGGTGCTGTGAATCGCGCCGATGCCCTGACCCGACCGC 720  
 QY 721 TGATCGCGATACCGTCCGCTGCTGCTTCGCGCTGAACGCGGACTTCAAGGCTT 780  
 DB 721 TGATCGCGATACCGTCCGCTGCTGCTTCGCGCTGAACGCGGACTTCAAGGCTT 780

Db 721 TGATCGCGATACCGTCCGCTGCTGCTTCGCGCTGAACAAAGCCGAGACTTCAAGGCTT 780  
 QY 781 TCGGCTGTATCGAGGAGAGCGCAGAAAGTGGTGGCTTCCAGTTTCGACAAAGCTGCTGC 840  
 Db 781 TCGGCTGTATCGAGGAGAGCGCAGAAAGTGGTGGCTTCCAGTTTCGACAAAGCTGCTGC 840  
 QY 841 CGGTTGGCGAACTGAAGATCCGTTGGCGCCACAACTATTTCACAGCGCTCGCGCGCTGG 900  
 Db 841 CGGTTGGCGAACTGAAGATCCGTTGGCGCCACAACTATTTCACAGCGCTCGCGCGCTGG 900  
 QY 901 CGTGGGCGATCGGTCGGCTCGCTTCGACGCCATGCTCGCGCGCTGAAGCGCTTTT 960  
 Db 901 CGTGGGCGATCGGTCGGCTCGCTTCGACGCCATGCTCGCGCGCTGAAGCGCTTTT 960  
 QY 961 CCGGCTGGCTCATCGCTGCGCTGCGGAGTACGCGAGCGGAGCGGCTGAGCTACTACGACG 1020  
 Db 961 CCGGCTGGCTCATCGCTGCGGAGTACGCGAGCGGAGCGGCTGAGCTACTACGACG 1020  
 QY 1021 ATTCCAAGGCCACCAACGTCGCGCGCCCTTGGCGCGCATCGAGGGGCTGGTGGCGACA 1080  
 Db 1021 ATTCCAAGGCCACCAACGTCGCGCGCCCTTGGCGCGCATCGAGGGGCTGGTGGCGACA 1080  
 QY 1081 TCGACGGCAAGCTGCTGCTCGCGCGGAGAGCGGCAAGGCGCGGATTTCCATGACG 1140  
 Db 1081 TCGACGGCAAGCTGCTGCTCGCGCGGAGAGCGGCAAGGCGCGGATTTCCATGACG 1140  
 QY 1141 TGGCGAGCGCGTTCGCGCGCTTCTGCGCGCGGTGCTGCTTGGCGGTGAGCGCGGGC 1200  
 Db 1141 TGGCGAGCGCGTTCGCGCGCTTCTGCGCGCGGTGCTGCTTGGCGGTGAGCGCGGGC 1200  
 QY 1201 TGATTCGCCAGGCACTGGGCAACCGGTACCGGTGCTGCGCGCTCGCAACGCTGGACGAG 1260  
 Db 1201 TGATTCGCCAGGCACTGGGCAACCGGTACCGGTGCTGCGCGCTCGCAACGCTGGACGAG 1260  
 QY 1261 CAGTCGCGAGCGCGCGGAGCTGGCGCGGAGAGCGGATGCGGTGCTGCTGCGCGGCT 1320  
 Db 1261 CAGTCGCGAGCGCGCGGAGCTGGCGCGGAGAGCGGATGCGGTGCTGCTGCGCGGCT 1320  
 QY 1321 GCGCGAGCTGGACATGTTCAAGAACTTCGAAAGCGGAGCGGCTGTTGCGCAAAAGCCG 1380  
 Db 1321 GCGCGAGCTGGACATGTTCAAGAACTTCGAAAGCGGAGCGGCTGTTGCGCAAAAGCCG 1380  
 QY 1381 TAGAGGAGCTAGCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1440  
 Db 1381 TAGAGGAGCTAGCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1440  
 QY 1441 GCACGGCATC 1450  
 Db 1441 GCACGGCATC 1450

RESULT 2

AAS54275

ID AAS54275 standard; DNA; 1347 BP.

XX AAS54275;

XX AC

XX 13-FEB-2002 (first entry)

XX DE

XX Pseudomonas aeruginosa DNA for cellular proliferation protein #406.

XX Antisense; ds; prokaryotic cellular proliferation gene;

XX Antibiotic; antibacterial; drug design.

XX OS

XX Pseudomonas aeruginosa.

XX WO200170955-A2.

XX PD

XX 27-SEP-2001.

XX 21-MAR-2001; 2001WO-US09180.

XX 21-MAR-2000; 2000US-191078P.



XX Salmonella typhi DNA for cellular proliferation protein #374.  
 XX DE  
 XX KW  
 XX Antisense; ds; prokaryotic cellular proliferation gene;  
 XX antibiotic; antibacterial; drug design.  
 XX OS  
 XX Salmonella typhi.  
 XX WO200170955-A2.  
 XX PD 27-SEP-2001.  
 XX PF 21-MAR-2001; 2001WO-US09180.  
 XX PR 21-MAR-2000; 2000US-191078P.  
 XX PR 23-MAY-2000; 2000US-206848P.  
 XX PR 26-MAY-2000; 2000US-207727P.  
 XX PR 23-OCT-2000; 2000US-242578P.  
 XX PR 27-NOV-2000; 2000US-253625P.  
 XX PR 22-DEC-2000; 2000US-257931P.  
 XX PR 16-FEB-2001; 2001US-269308P.  
 XX PA (ELIT-) ELITRA PHARM INC.  
 XX PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GU;  
 XX PI Yamamoto RT, Xu HH;  
 XX PR WPI: 2001-611495/70.  
 XX PR P-PSDB; AAU38482.  
 XX DR New polynucleotides for the identification and development of  
 XX PT antibiotics, comprise sequences of antisense nucleic acids -  
 XX PS Claim 27; Seq ID No 9978; 511pp; English.  
 XX CC The invention relates to antisense inhibitors of genes essential to  
 CC prokaryotic cellular proliferation, their use in identifying the  
 CC genes, their use in the discovery of novel antibiotics, the essential  
 CC genes, themselves and the encoded proteins. The prokaryotes used are  
 CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella  
 CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The  
 CC invention is also useful for the identification of potential new targets  
 CC for antibiotic development. The antisense nucleic acids can also be used  
 CC to identify proteins used in proliferation, to express these proteins,  
 CC and to obtain antibodies capable of binding to the expressed proteins.  
 CC The proteins can be used to screen compounds in rational drug discovery  
 CC programmes. The antisense nucleic acid sequence is also useful to screen  
 CC for homologous nucleic acids which are required for cell proliferation in  
 CC a wide variety of organisms. The present sequence encodes an  
 CC essential prokaryotic cellular proliferation protein.  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 XX SQ Sequence 1317 BP; 258 A; 341 C; 427 G; 291 T; 0 other;

Query Match 23.0%; Score 332.8; DB 23; Length 1317;  
 Best Local Similarity 56.1%; Pred. No. 3.6e-51;  
 Matches 747; Conservative 0; Mismatches 552; Indels 33; Gaps 5;

QY 69 GACCACTTCGGATCTTGGCGCTCGGCGAGAGCGGATGTCCTGGTGGCGTACCTG 128  
 DB 16 GACAAAAGCTGCTATTATCGTCTGGGCTTAACCGGACTCTCTTGGCTGGACTTTT 75  
 QY 129 GCGCGCGCGGCTTGGCTTCGCGGTGTCGATACCCGAGAACCCCGGAGCTGGCC 188  
 DB 76 CTCGCCCGGGGTGACGCCGGGGTGGAGTTCGCTTGGCGGAGTACCTGACGCCCGGCTGGAT 135  
 QY 189 ACCCTGCGTGCCAGTATCCGAGGTGGAGTTCGCTTGGCGGAGTACCTGACGCCCGAGTTC 248  
 DB 136 AAGTGGCG-----CAAGAGTTGAGCGTCAAGTGGCGGCTGAGAACCGAGTGG 186

QY 249 CTCGTCTCGCGCGCGAACTCTATGTCAGCCCGCGCTTGTGCTGCTGCGACCCCTGGCGTG 308  
 DB 187 CTCCTAGCGGGGATTAATCGTCGCCAGCCCTGGTATGCTCTGCGGATCCCTCTCTT 246  
 QY 309 GTACAGCGCGCGCGAAAGCGTGCATCTCCGGTGACATCATCTCTTCGCCCCCGGAG 368  
 DB 247 AGCGCTGCGCGCGAGCGCTGGAGTGAATTTGTCGGCATATCGAACTGTTTTCGCGGAA 306  
 QY 369 GCGAAGCGCGCGATCGTCGCCATCAGCGTTCCAAACGCGAAGACACCGTGACACCCCTG 428  
 DB 307 GCGCAAGCGCGGATTTGTGGCCATCACCGGCTCGAACGCGAAAGACCGGTGACCACTTA 366  
 QY 429 GTGGCGAAATCGCGTGGCGGACAAAGCGTGTGCGGCTCGCGCGCAACCTTCGCGACC 488  
 DB 367 GTGGCGAGATGCGGAAGCGGGCGGCGTCAATGTGCGGCTTGGCGCAATATCGCTCT 426  
 QY 489 CCGGCGCTCGACTGCTGGCGGACGACATCGAGCTGTAGTGTGGAGCTGTGCGAGCTTC 548  
 DB 427 CCGGCGCTGATGCTGTGATGCCGATCGCAATTTAGCTGTGTGAATTTATCCAGTTTC 486  
 QY 549 CAGCTGGAACCTGCGATCGCTCAACGCGGAGTGGCGACCGTGTGTAACGTCACGCGAA 608  
 DB 487 CAATTGGAACCTACCTCAAGTTTGAAGCGGCGCGCAACGGTGTCAACGTCACGTAA 546  
 QY 609 GACCATATGATCGCTA---CGACGCGATGGCTGACTACCACTTGCGCCAAAGCACCGGATC 665  
 DB 547 GATCATATGACCGGTATCCGTTTGGTTTGAACAGTACCGGGCGCGAACTGCGCGTC 606  
 QY 666 TTCCCGGCTGCGCGCGAGTGTGGTGAATCGCGCGATGCGCTGACCGCGAGCTGATC 725  
 DB 607 TAGGAGAGCGGAAAGTGTGCTAGTGAATGCGGATGACGCTGACGATGCGCGTACGC 666  
 QY 726 GCGCATACCGTCCGCTGCTGCTGGGCTGAACAAGCGCGACTTCAAGGCTTTCGCG 785  
 DB 667 GGGCGCGATGAGCGCTGCTGCTGCTTGGCTCAATATGGTGTATATCA-----C 717  
 QY 786 CTGATCGAGAAAGCGCGAGAGTGGTGGCGCTTCCAGTTCGACAAGTGTGTCGGT 845  
 DB 718 CTTAATCGTCAGCAGGGCGAAACCTGGCTACGAGTCAAAAGGTGAGAAGTGTGTAACGTG 777  
 QY 846 GCGCACTCAAGATCGTGGCGCGCCCACTTCCAAACGCGCTCGCGCGTGGCGCTG 905  
 DB 778 AAGAGATGAAGCTTCCCGTCAAGTAACTATACCAATGCGTTACGGCGCTGGCGCTG 837  
 QY 906 GCGCATGCGGTGCGCTGCTGCGCGCATGCTGCGCGCGCTGAAGCGTTCCTCCGCG 965  
 DB 838 GCGGATGCGGTAGTCTGCGCGCGCGCAGCTTTGAAGCGCTTGACGACATTTACCGGT 897  
 QY 966 CTGGCTCATCGTGGCGAGTGGGTACGAGCGGCGAGGCGGTGAGCTACTACGACGATTC 1025  
 DB 898 CTGGCGCATCGCTTCAGCTGGCGTGGAGCATAAACGGGCTTCGTTGGATCAACGACTCG 957  
 QY 1026 AAGGCCCAACGTCGGCGCGCCCTGCGCGCGATCGAGGGCTGGGTCCGCGACATCGAC 1085  
 DB 958 AAGCGCAACATGTGCGCAGTACCGAAGCGCGCTAAACGGTTTGCATG-----TGGAC 1011  
 QY 1086 GCGAAGCTGCTGCTGCTCGCGCGGAGACGCGCGCGATTTCCATGACCTCGCGC 1145  
 DB 1012 GTACGCTCATCTGCTGCTGCGCGCGAGCGTAACTGCGCAGACTTTTTCCTCGCTGGCG 1071  
 QY 1146 GAGCGGTGCGCGCTTCTGCGCGGGGTGCTGCTGCTGGCGGTGACCGCGCGGTGAT 1205  
 DB 1072 CGCTATCTGACCGGGATCGTATCGCGCTGTATGCTTGGCGGGATGGCGCGAGCTT 1131  
 QY 1206 GCGCAGGCACTGGCAACGCGGTACCGCTGCTGCGCTGCAACGCTGACCAAGCAGTC 1265  
 DB 1132 GCGGCACTG-----GTCCGGAATTCGCCCAACAGACTGACAGATGAAGAGGCGATG 1185  
 QY 1266 GCGCAGGCGCGGAGCTGGCGCGCGGAGAGCGATGGGTGCTGTTGTCGCGCGCTGCGCG 1325  
 DB 1186 CGTTTGTGCGCGCGGCTTACGCGGGGTGATGTTGCTGTTGTCGCGCGCTGCGCG 1245  
 QY 1326 AGCTGGACATGTTCAAGAACTTCGAAGAACCGCGGACGCTTTCGCCCAAGACCGTAGAG 1385





QY 977 CTGCCAGTGGGTACCGGAGCGCGAGCGGTGAGCTACTACGACGATTCCAAAGCCACAA 1036  
DB 909 CTTTGAAGTGTGTGGAGCATACGCGGTACGTTGGATTAAAGTTCGAAAGCGACCA 968  
QY 1037 CGTGGCGCGCCCTGGCGGGGATCGAGGGGTGGTCCGACATCGAGCGCAAGCTGGT 1096  
DB 969 CGTGGCAGTACGGAAGCGGCGTGAATGGCTG-----CACGTAGACGGCACACTGCA 1022  
QY 1097 GCTGCTCGCGCGGAGCGGAGCGGCGGATTTCCATGACCTGCGGAGCGCGTGGC 1156  
DB 1023 TTTGTTGCTGGTGGCGATGTAATCGCGGACTTTAGCCACTGGCGGCTTACCTGAA 1082  
QY 1157 GCGCTTCGCCGGCGGTGCTGTTGGCGGTGACCGCGGCTGATTGCCAGGCACT 1216  
DB 1083 TGGCGATAACGTACGCTCTGTTATTTGTTGCTGCGTACGCGCGCGAGCTGGCGGCTAC- 1141  
QY 1217 GGGCAACCGGTACCGTGTGGCGGTGCGAAGCTGGACGAGCAAGTGGCGGCGCGC 1276  
DB 1142 -----GCCCGGAAGTGGCAGAACAAACGAACTATGGAACAGCGGATGCGCTTGTCTGGC 1196  
QY 1277 CGAGCTGGCGCGGAAGCGGATGCGGTGCTGTTGTCGCGGCTGCGCGAGCGTGGACAT 1336  
DB 1197 TCCGCGTGTTCAGCGGCGGATATGTTCTGCTCTCCCGCGCTGTGCCAGCTTGATCA 1256  
QY 1337 GTTCAAGAACTTCGAAGAACCGCGGAGCGCTGTTCGCCAAAGCCGTAGAGGAGTACGCTG 1396  
DB 1257 GTTCAAGAACTTTGAACAACGAGGCAATGAGTTTCCCGTCTGCGGAAGGAGTTAGGTTG 1316  
QY 1397 A 1397  
DB 1317 A 1317

RESULT 6  
AAZ53175/C  
ID AAZ53175 standard; DNA; 1380 BP.  
XX  
AC AAZ53175;  
XX  
DT 21-MAR-2000 (first entry)  
XX  
DE Neisseria meningitidis ORF 090 partial DNA sequence SEQ ID NO:301.  
XX  
KW Neisseria meningitidis; Neisseria gonorrhea; antigen; vaccine;  
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;  
KW antibacterial; gene therapy; ds.  
XX  
OS Neisseria meningitidis.  
XX  
PN W09957280-A2.  
XX  
PD 11-NOV-1999.  
XX  
PF 30-APR-1999; 99WO-US09346.  
XX  
PR 01-MAY-1998; 98US-0083758.  
PR 31-JUL-1998; 98US-0094869.  
PR 02-SEP-1998; 98US-0098994.  
PR 02-SEP-1998; 98US-0099062.  
PR 09-OCT-1998; 98US-0103749.  
PR 09-OCT-1998; 98US-0103794.  
PR 09-OCT-1998; 98US-0103796.  
PR 25-FEB-1999; 98US-0121528.  
XX  
PA (CHIR ) CHIRON CORP.  
PA (GENO-) INST GENOMIC RES.  
XX  
PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;  
PI Petersen J, Pizsa M, Rappuoli R, Ratti G, Scalato E, Scarselli M;  
PI Tettelin H, Venter JC;  
XX  
DR WPI: 2000-062150/05.  
DR P-PSDB; AAY74413.

XX Novel Neisserial polypeptides predicted to be useful antigens for  
PT vaccines and diagnostics  
PS Claim 7; Page 286; 1453pp; English.  
XX  
CC AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941  
CC represent novel Neisseria meningitis and N. gonorrhea polynucleotides  
CC and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ54573 represent  
CC PCR primers used in the exemplification of the present invention. The  
CC polypeptides, the polynucleotides, antibodies and compositions of  
CC the invention can be used as vaccines, as diagnostic reagents, and as  
CC immunogenic compositions. The polypeptides can be used in the  
CC manufacture of medicaments for treating or preventing infection due to  
CC Neisserial bacteria (e.g. meningitis and septicaemia), to detect the  
CC presence of Neisseria bacteria, or to raise antibodies. They may also  
CC be used to screen for agonists or antagonists, which may themselves  
CC have use as antibacterial agents. The polynucleotides of the invention  
CC may also be used in gene therapy protocols.  
XX  
SQ Sequence 1380 BP; 292 A; 424 C; 346 G; 318 T; 0 other;  
Query Match 12.4%; Score 180; DB 21; Length 1380;  
Best Local Similarity 49.7%; Pred. No. 1.1e-23;  
Matches 662; Conservative 0; Mismatches 620; Indels 50; Gaps 6;  
QY 80 CATCGTTGTCGGCCCTCGCAAGAGCGGATGTCCTGTGCTGCTACCTGCGCGCGCGG 139  
DB 1342 CTTGTCGCGGACTCGCGGTACGGGTATTTCCATGATGCTTACCTGCGCAAAACGG 1283  
QY 140 CTTGCTTTTCGCGGTGTCGATACCCGAGAGAACCCCGGAGCTGGCCACCTCGGTGC 199  
DB 1282 CGCGGAGTGTGTCGCTATGATGCGGAGCTGAAGCCGGAACGCGTGTGCAAAATCGGTAA 1223  
QY 200 CCAGTATCCGAGGTGGAGTCCGTTGCGGCACTCGACCGCGAGTTCCTGCTCGCGC 259  
DB 1222 GATGTTGACGGTGTGGTGTTTTACACGGGCGCTGTAAGATGCGTGGACACGGTTT 1163  
QY 260 CCGCAACTCTATGTCAGCCCGCGCTTGTGCTGCGCACCCCTGCGCTGGTACAGCGCGC 319  
DB 1162 CGATATTCGCTCTCAGTCCCGGATCAGCGAGCGCGGATATGAGCGCTTCAA 1103  
QY 320 CGCGAAGCGGTGCGCATCTCGGTGACATCATCTTTCGCCCGCGAGCGGCAAGCGGCC 379  
DB 1102 GCAAAACGCGAGCGGTGTTGGCGACATCAATTTGCTGGCGGACATTTGTAACCGCG 1043  
QY 380 GATCGTC-----GCCATCACCGTTCACCGAGAGACCGGTGACCACTCGT 430  
DB 1042 GCACGACAGGTAAATTGCGATTACCGGAGCAACCGGCAAAACCGGTAAACGAGCTGT 983  
QY 431 GGGCGAAATGGCGTGGCGCGGACAAAGCTGTCGCCGTCGCGGCAACCTCGGACCC 490  
DB 982 CGGCTATCTCTATCAAGTGGCGGTGATACCGTTATCGGGGCAATATCGGACGCC 923  
QY 491 GCGGCTCG-----ACCTGCTGCCCGACGACATFCGAGCTGTAGTGTGGAGCT 538  
DB 922 GGTGTTGGAGCGGAATGCGACGCGCAAGGCAAAAGCGGACGTGTGGGTGTTGAGCT 863  
QY 539 GTCGAGCTTCCAGCTGGAAACCTGCGATCGCTCAACCGCGAGGTGGCGACCGTCTGAA 598  
DB 862 TTCCAGCTTCCAACTGGAAACACCGGAAGCTGGCTCGGACTGGCGGACGCTGCTGAA 803  
QY 599 COTCAGCGAAGACCATATGATGCTGCTAGCAGCGCATGCTGACTACCACTGCCCAAGCA 658  
DB 802 CATTTCCGAGACCATCTCGACCGCTACGACGATTTGCTGACTATGCGCATACCAAGC 743  
QY 659 CGGATCTTCGCGGTGCCCGCAGGTGTGATCGCGCGCGATGCGCTGACCGGAC 718  
DB 742 CAAGATTTTCCGTGGCGACGCGTGCAGGTTTGAATGCGGACGATGCTGCTGCGCGC 683  
QY 719 GCTG-ATCGCGCATACCGTGGCTGCTGCTGCGCTGACAAAGCGGACCTCAAG 777  
DB 682 GATGAAGCTGCCGGCGCGAGGTAATGTTTTCGTTGGAACAGGAGCTGATTTCTG 623







```
Db 52199 GCTTTTGGAGGCGAATGCGACGCGAAGGCAAAAGGCGGACGTGTGGGTGTTGAGCT 52140
QY 539 GTCGAGCTTCCAGCTGGAACCTGCGATCGCTCAACGCCGAGGTGGCGACCTGCTGAA 598
Db 52139 TTCCAGCTTCCAACTGGAACACCAAGCGCTCGGTCCGACGTGCGGCGGAGGCTGCTGAA 52080
QY 599 CGTCAGCGAAGACCATATGGAATGCTAGCAGCGCATGCTGACTACACCTGGCCAAAGCA 658
Db 52079 CATTTCCGAGACCATTCGACCGCTACGACGATTCGCTGACTATGCGCATACCAAGC 52020
QY 659 CCGGATCTTCCCGGTGCCCGCCAGGTGCTGTGTAATCGCGCCGATGCGCCCTGACCCGACC 718
Db 52019 CAAGATTTTCCGTGCGCAGCGCGTGCAGGTTTGAATGCGGACGATGCGTTCGCGCGC 51960
QY 719 GCTG-ATCGCGCATACCGTCCGCTGCTGCTCGGCTGCAACAGCGCGGACTTCAAG 777
Db 51959 GATGAAGCGTCCGCGCGGAGGTAAATGTTTTCGTTGGAAACAGAGCTGATTTCTG 51900
QY 778 CTTTCGGCTGATCGAGGAAGACGCGCAAGTGGCTGGCGTTCAGTTTCGACAAAGCTGC 837
Db 51899 GTTGAACGCGAGACGCGCGCTGAACAAAGGCAATGAAG-----ATTGGA-51853
QY 838 TCCCGGTTGGCAACTGAAGATCCGTGGCGGCCCAACTATTTCCAAAGCGCTCGCGCGC 897
Db 51852 TTGTCAAGCAAGACATTCGTTGCAAGGTCTGCACACGCGCGCTAACGTCATGCTGCGC 51793
QY 898 TGGCGCTGGCGCATCGGTGCGCTGCGCTTCGACGCCATGCTCGCGCGCTGGAAGGCT 957
Db 51792 TGGCTTTGTGAGGCCATCGTTTGTGCGCGAAGCATGCTCAACACGTCAAACCT 51733
QY 958 TTTCCGGCTGCTCATGCTGCCAGTGGGTACGCGAGCGGCGGCTGAGCTACTAG 1017
Db 51732 TCCAAAGGCTCGCGCACCGCTGGGAAATAATCGCGGAGAAAACGCGTGGTGTATTCG 51673
QY 1018 AGATTCGAAGCCACCACTGCGCGCGCGCTGCGCGCGATCGAGGGCTGGGTGCGC 1077
Db 51672 ACGACAGCAAGGCGACGATGTCGCGCGGACTGCGCGCGGAT-----TGCGG 51625
QY 1078 ACATCGACGCGAAGCTGCTGCTGCTCGCGCGGAGCGCAAGGCGCGGATTTCCATG 1137
Db 51624 GTTTCGAATCCGCTCTCGTGATTTTGGCGGCGATGGTTAAGGCGAGGACTCACGC 51565
QY 1138 ACTCGCGAGCGGTGCGCGCTCTGTCGCGCGCGGTGTACTGCTTGGCCGTGACGCGG 1197
Db 51564 CCGTCGCGATCACTGTGAGCAAGGCAAAAGGCGTGTCTTGTATGCTGCTGATGCGC 51505
QY 1198 GCTGATTCGCCAGGCACTG---GGCAAGCGGTACCGCTGCTGCGCGCTCGCAAGCTGG 1254
Db 51504 CGCAATCCGCGGATTTGGAGCGCTGCGGCTTGAATATGACCGACTGCGCCACTTTGG 51445
QY 1255 ACGAAGCACTCGCGCAGCGCGCGAGCTGGCCCGGAGCGGATGCGTGTGTTGTCG 1314
Db 51444 GAGAAGCGTTTCAGACGGCATATGCCAAGCGCAAGCAGCGCATATTGTTGCTCAGCG 51365
QY 1315 CGCGCTCGCGAGCTGGACATGTCAGAACTTCAGAAACGCGGACCGCTGTTCGCCA 1374
Db 51384 CGCGCTCGCGAGCTTGTATGTTTCAAGGCTACGCGCACCGTTCGGAAGTGTATCG 51325
QY 1375 AAGCGGTAGAGG 1386
Db 51324 AAGCGTTAAGG 51313
```

## RESULT 8

AAF21607

ID AAF21607 standard; DNA; 349980 BP.

XX

AC

XX

DF

XX

DE

```
XX Neisseria meningitidis; Neisseria gonorrhoeae; immunogenic; vaccine;
KW diagnosis; antigen; detection; infection; gene therapy; antibacterial;
KW ds.
XX Neisseria meningitidis.
XX WO2000066791-A1.
XX 09-NOV-2000.
XX 08-MAR-2000; 2000WO-US05928.
XX 30-APR-1999; 99US-0132068.
XX 08-OCT-1999; 99WO-US23573.
XX 28-FEB-2000; 2000GB-0004695.
XX (CHIR ) CHIRON CORP.
XX (GENO-) INST GENOMIC RES.
XX Pizza M, Hickey E, Peterson J, Tettelin H, Venter JC, Massignani V;
XX Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V, Rappuoli R;
XX Frazer CM, Grandi G;
XX WPI; 2000-647603/62.
XX Neisseria meningitidis B full length genome sequence and open reading
frames are used to detect, treat and prevent Neisserial infections .
XX Claim 7: Appendix A; 692pp; English.
XX The present invention describes the full length genome of
Neisseria meningitidis B (NMB). The sequences in AAF21544 and AAF21607
to AAF21613 represent fragments of the NMB genomic sequence, as the
sequence was too long to go in a record on its own it was split into 8
sequences which overlap each other at the beginning and end of each
sequence by 49980 bp (i.e. the last 49980 bp of AAF21544 is repeated at
the beginning of AAF21607, the last 49980 bp of AAF21607 are repeated at
the beginning of AAF21608, and so on). AAF21545 to AAF21588 encode the
Neisseria proteins given in AAB58550 to AAB58593, and AAF21589 to
AAF21606 represent PCR primers which are used in the exemplification of
the present invention. The NMB genome and fragments from it have
antibacterial activity, and can be used in vaccines and gene therapy.
Neisseria nucleic acids, proteins and/or antibodies which binds to the
proteins can be used in compositions for treating or preventing infection
due to Neisserial bacteria or as a diagnostic reagent for detecting the
presence of Neisserial bacteria or of antibodies raised to Neisserial
bacteria. Computers, computer memory, computer storage medium or computer
databases can be used in a search to identify open reading frames (ORFs)
or coding sequences within the NMB genome. The DNA sequences provide
further opportunities to find antigenic or immunogenic proteins which are
more effective in vaccines than the outer membrane proteins currently
used.
```

Sequence 349980 BP; 84410 A; 84863 C; 94187 G; 86520 T; 0 other;

Query Match 12.4%; Score 180; DB 21; Length 349980;  
Best Local Similarity 49.7%; Pred. No. 1.1e-23;  
Matches 662; Conservative 0; Mismatches 620; Indels 50; Gaps 6;

```
QY 80 CATCGTTTCGCGCTCGGCAAGAGCGGATGTCCTGTGGTGGCTACCTGGCGGCGCGCG 139
Db 128568 CCTCGTCGCGGACTCGCGGTATTTCCATGATTGCTGCGCAAAAACG 128627
QY 140 CTTGCTTTTCGCGTGTGATACCGGAGAACCGCGGAGCTGGCCACCTCGGTGC 199
Db 128628 CCGGAGGTGCTGCGGTATGATGCGGAGCTGAAGCGGACGCGTGCAGAAATCGGTAA 128687
QY 200 CCAGTATCCGAGGTGGAAGTTCGTCGCGGAACTCGACGCGGAGTTCCTGTGCTCCGC 259
Db 128688 GATGTTGACGGTGTGCTGTTTACACGCGCGCTCTGAAGATGCGTGGACACAGCTTT 128747
QY 260 CCGCGAACTCTATGTAGCCCGCGCTTGTGCTGCGCACCCCTGCGCTGTACAGCCCGC 319
```

129803 CGCGTCGCGAGCTTTGATATGTTCAAAGGCTACGCGACGTTTCGGAAGTGTATTACG 129866

1375 AAGCCGTAGAGG 1386  
 |||||

129863 AAGCGTTTAAGG 129874

RESULT 9  
 AAA81490  
 ID AAA81490 standard; DNA; 1437668 BP.

XX  
 AC AAA81490;  
 XX  
 DT 04-DEC-2000 (first entry)  
 XX  
 DE N. meningitidis B full length genome DNA sequence SEQ ID NO:1068.  
 XX  
 KW Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;  
 KW antigen; vaccine; diagnosis; infection; antibacterial; identification;  
 KW Meningococcus B; MenB; ds.  
 XX  
 OS Neisseria meningitidis.  
 XX  
 OS WO200022430-A2.  
 XX  
 PD 20-APR-2000.  
 XX  
 PF 08-OCT-1999; 99WO-US23573.  
 XX  
 PF 09-OCT-1998; 98US-0103794.  
 XX  
 PR 30-APR-1999; 99US-0132068.  
 XX  
 XX (CHIR ) CHIRON CORP.  
 XX  
 PI Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;  
 PI Massignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;  
 PI Rappuoli R, Pizza M;  
 XX  
 XX WPI: 2000-318079/27.  
 XX  
 XX Isolated nucleotide sequences of Neisseria meningitidis which can be  
 XX used in the diagnosis and treatment of N. meningitidis infection and  
 XX other Neisserial infections, for example, N.gonorrhoea -  
 XX  
 XX Claim 7: Page 866-1272; 1760pp: English.

The present invention describes methods of obtaining immunogenic proteins from Neisseria genomic sequences. AAA81453 to AAA82414 represent specifically claimed Neisseria meningitidis genomic DNA sequences; AAA81260 to AAB25620 to AAB25663 represent Neisseria DNA sequences and their corresponding proteins; AAA81254 to AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the isolation of Neisseria meningitidis DNA sequences; and AAA81322 to AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF sequences, which are all used in the exemplification of the present invention. The nucleic acid sequences, protein sequences, and antibodies against them, can be used in the manufacture of a composition. The composition can be used as a medicament (or in the manufacture of a medicament) for treating, preventing or diagnosing infection due to Neisserial bacteria. For example, some of the identified proteins could be components of vaccines against Meningococcus B; against all serotypes; and/or against all pathogenic Neisseriae. Identification of sequences from the bacterium will also facilitate production of biological probes, particularly organism-specific probes. Attempts to make efficacious Meningococcus B vaccines have failed mainly due to antigen tolerance. Multivalent vaccines have also been tried but none have successfully overcome antigenic variability. The provision of further, complete sequences may provide an opportunity to identify secreted or surface exposed proteins that may be presumed targets for the immune system and which are not antigenically variable or at least more conserved than other more variable regions.

Sequence 1437668 BP: 344338 A: 353206 C: 385074 G: 355045 T:5 other;  
 XX  
 XX

Query Match	12.4%	Score 180;	DB 21;	Length 1437668;
Best Local Similarity	49.7%;	Pred. No. 1.1e-23;		
Matches 662;	Conservative 0;	Mismatches 620;	Indels 50;	Gaps 6;
Qy	80	CATCGTTGTTCGGCTCGGCAAGACGGCATGTCCTGTGTGCGCTACCTGGCGCGCGCGG	139	
Db	428568	CCTCGTCGGGCACTCGGCGGTACGGGTATTTCCATGATGCCYACCTGCGCAAAACGG	428627	
Qy	140	CTTGCCCTTTCCCGTGTGCGATACCCGAGAACCCGCGGAGCTGGCCACCTCGGTGC	199	
Db	428628	CGCGAGGTTCTCGTATGATCGGAGCTGAAGCCGGAACGCGTGTGCAAAATCGGTAA	428687	
Qy	200	CCATATCCCGCAGTGGGAAGTCGTTGCGGGGAACTCGACGCCGAGTTCTCTGCTCCGC	259	
Db	428688	GATGTTGACGGTGTGGTGTTTACACGGCGCGTCTGAAGATGCGCTGGACAACGTTTT	428747	
Qy	260	CCGCGAACTCTATGTCAGCCCCGGCTTGTGCTGGCACCCCTGCGCTGGTACAGGCCGC	319	
Db	428748	CGATATTCTGTGCTCTCAGTCCCGCATCAGCGAGCGGACCGCGATATTGAGGCGTCAA	428807	
Qy	320	CGCGAAAGCGTGCSCATCTCCGGTGACATCGATCTCTCGCCCGGAGGCGAAGCCCC	379	
Db	428808	GCAAAACGGCGGACCGGTGTTGGGGGCATCGAATGCTGGCGGACATGTGAACCGCGG	428867	
Qy	380	GATCGTC-----GCCATCACCGGTTCCAACGCGAAGACACCGTGACCACTCTGT	430	
Db	428868	GGACGACAAGTAATTGCGATTACGGCGACGAACGCAAAACACCGGTACAGCCTGT	428927	
Qy	431	GGGCGAAATGGCGGTGGCCGGGACACGCTGTGCCGTGGCGGGCACTCGGCACCCC	490	
Db	428928	CGGTATCTCTGTATCAAGTCGGGCTGCGATACCTTTATCGGGCAATATCGCACGCC	428987	
Qy	491	GGCGCTCG-----ACCTGCTGGCGACGACATCGAGCTGTAGCTGTTGAGCT	538	
Db	428988	GGTTTGGAGCGGAAATGGCAGCGCGCAAGGCAAAAGGCGGACGTGGGTGTTGAGCT	429047	
Qy	539	GTCGAGCTTCCAGTGGAAACCTGCGATCGCTCAACGCCGAGGTGGCGACCGTCTGAA	598	
Db	429048	TTCCAGGTTTCCAACGTGAAAAACACCGAAAGCTGCGTCCGACTGCGGGCACGCTGTA	429107	
Qy	599	GGTCAGCGAAGACCATATGGATGCTACGAGGCGATGCTGACTACCACCTGGCCAAAGCA	658	
Db	429108	CATTCCGAAGACCATCTCGACCGCTACGAGACCTGTCTGACTATGGCGATACCAAAGC	429167	
Qy	659	CCGATCTTCCGCGGTGCCGCCAGGTGCTGGTGAATCGCGCGATGCCCTGACCGCAC	718	
Db	429168	CAAGATTTCCGTGGCAGCGCTGCAGTTTGAATCGGACGATGCGTTCTGCGCGC	429227	
Qy	719	GCTG-ATCGCCGATACGTCGCGTGTGCTGTTTCGGCCTGAACAAGCGGACTTCAAGG	777	
Db	429228	GATGAAGCGTCCGGCGCGAGGTAAATGTTTTCTGTTGGAACACGAAGCTGATTTCTG	429287	
Qy	778	CTTTTCGGCCTGATCGAGGAACAGCGCCAAAGTGGCTGGCGTTCCAGTTTCGACAAAGCTC	837	
Db	429288	GTTGGAACGCGAGACAGCGCCCTGGAACAGGCAATGAAG-----ATTGA	429334	
Qy	838	TGCGGTTGGCGAACTGAAGATCCGTGGCGGCCACAACATATPCCAACGCGCTCGCGCGC	897	
Db	429335	TTGTCAAGCAAGACATTCGGTTGCAAGGCTCGACAACGCCGCTAACGTCATGGTGC	429394	
Qy	898	TGGCGCTGGCCATGCGGTGCGCTGCGTTTCGAGCCCATGCTCGCGCGCTGAAGCGT	957	
Db	429395	TGGCTTGTGTGAGGCCATCCGTTGTTCGCGGAAGCATGTCTCGAACACGTCFAAAACCT	429454	
Qy	958	TTTCCGGCTGGCTCATCGCTGCCAGTGGGTACGCGAGCGGAGGCGCTGAGCTACTAGC	1017	
Db	429455	TCCAAGGCTCCCGACCGCTGGAAAAATCGGCGAGAAAAACGCGTGTGTTTATCG	429514	
Qy	1018	ACGATTTCAAGCCACCAACGTTCGCGCCGCCCTCGCGCGCGATCGAGGGGCTGGGTGCGG	1077	
Db	429515	ACGACAGAAAGGACGAATGTCGCGCGATTCGCGCGCATTCGCGCGCAT-----TGCGG	429562	

RESULT 10	
AAH52025	
ID	AAH52025 standard; DNA; 1461 BP.
XX	
AC	AAH52025;
XX	
DT	04-SEP-2001 (first entry)
XX	
DE	Mycobacterium tuberculosis poten
XX	
KW	Drug target; growth; organism vir
XX	
OS	Mycobacterium tuberculosis.
XX	
PN	WC200135317-A1.
XX	
PD	17-MAY-2001.
XX	
PF	13-NOV-2000; 2000WO-US31152.
XX	
PR	12-NOV-1999; 99US-0165086.
XX	
PR	12-NOV-1999; 99US-0165124.
XX	
PR	01-FEB-2000; 2000US-0179531.
XX	
PA	(REGC ) UNIV CALIFORNIA.
XX	
PI	Eisenberg D, Rotstein SH, Marc
XX	
WPI	2001-329193/34.
DR	P-PSDB; AAG81174.
DR	

Identifying nucleotide or polypeptide sequence for use as drug target,  
involves providing algorithm that analyzes a functional relationship  
between nucleotide or polypeptide sequences, and comparing the  
sequences -  
Disclosure; Page 100;. 207pp; English.

This invention relates to a method for identifying a nucleotide or  
polypeptide sequence that may be a drug target, or essential for growth  
or viability of an organism. Polynucleotide sequences AAH51947 - AAH52092  
represent DNA encoding proteins AAG81096 - AAG81241, Mycobacterium  
tuberculosis proteins which are potential drug targets. The DNA and  
protein sequences are used to illustrate the method of the invention. The  
method involves providing an unknown nucleotide or polypeptide sequences,  
and comparing it to a number of sequences along with at least one  
algorithm capable of analysing a functional relationship between  
nucleotide and polypeptide sequences. The method is useful for

CC characterising the function of nucleic acids and polypeptides that may be  
 CC useful as a target for a drug or essential for the growth or viability of  
 CC an organism.  
 XX  
 SQ Sequence 1461 BP; 184 A; 472 C; 558 G; 247 T; 0 other;

Query Match 12.0%; Score 173.8; DB 22; Length 1461;  
 Best Local Similarity 51.7%; Pred. No. 1.5e-22;  
 Matches 499; Conservative 0; Mismatches 427; Indels 39; Gaps 3;  
 QY 273 GTACAGCCCGGCTGCTGCGGACACCCCTGCGTGTACAGGCGCGCGGAAGGGGTG 332  
 DB 214 GCCAGTCCCGCTTCTCGCGCAACCCCTACTGCGCGCGCGCGGGGGGTG 273  
 QY 333 CGCATCTCCGCTGACATCTCTTCCGCCG-----CGAGGCG 371  
 DB 274 CCGATCTGGGTGACGTGGAGTTAGCTGCGGGGTAGACCGAGGGGCTCTACGGACCG 333  
 QY 372 AAGGCCCGATCGTCCCATCACCGTTCCAGCGCAAGACACCGCTGACACCCCTGGTG 431  
 DB 334 CCGCGCAGCTGGCTGTGTGTGACCGGCAACCGCAAGACACCGAGCTCGATGCTG 393  
 QY 432 GCGGAATGCGGTGGCGCGGACACAGCTGTGCGCGTCCGCGCAACCTTCGCGACCCG 491  
 DB 394 CAGCCCATGTGATCGCGCGTGGCGCGCGCGCTGTGTGCGCAATATCGGCAGTGGC 453  
 QY 492 GCGTCGACCTGTGCGCGGACGACATCGAGCTGTACGTGTGGAGCTGTGAGCTTCCAG 551  
 DB 454 GTGCTGATGTGCTGGAGACCGCGCGAGCTGTGCGCGTGGAGTGTCCAGTTCCAG 513  
 QY 552 CTGGAACCTGCGATCCCTCAACGCGGAGGTGGCGACCGTGTGACGCTTCAGCAAGAC 611  
 DB 514 CTGCACTGGCGCGCTGCTGCGCGCGCGCGCGCGCTGTGCGCAATATCGGCAGTGGC 573  
 QY 612 CATATGATCGCTACGACGCGATGCTGACTACCTGCGCAAGCAGCGGATCTCCG 671  
 DB 574 CACCTGACCTGATGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 633  
 QY 672 GGTGCCCGCAGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 731  
 DB 634 GCGGGGTAGCGTGGCGCGCGTGGATGACAGCGCGCGCGCGCGCGCGCGCGCGCGTCA 693  
 QY 732 ACCGTGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 788  
 DB 694 CCGCGCAGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 753  
 QY 789 ATCAGGAAGACG 848  
 DB 754 CCG 813  
 QY 849 GAATGAGATCGCTGCG 908  
 DB 814 TCGATACCGTGGCG 873  
 QY 909 CATGCGGTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 968  
 DB 874 CGTCTGCTGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 933  
 QY 969 GCTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1028  
 DB 934 CGACACCG 993  
 QY 1029 GCCACCAACGCTGCG 1088  
 DB 994 GCCACCAACG 1038  
 QY 1089 AAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1148  
 DB 1039 AGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1098  
 QY 1149 CCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1208  
 DB 1099 GCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1158

QY 1209 CAGGC 1213  
 DB 1159 GAGGC 1163

RESULT 11  
 AAI99682/c  
 ID AAI99682 standard; DNA; 4411529 BP.

XX AAI99682;  
 XX 15-JAN-2002 (first entry)

XX Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 1.

XX Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;  
 KW variation; epidemiology; patient treatment; epidemic monitoring; ds.

XX Mycobacterium tuberculosis.

XX US6294328-B1.

XX 25-SEP-2001.

XX 24-JUN-1998; 98US-0103840.

XX 24-JUN-1998; 98US-0103840.

XX (GENO-) INST GENOMIC RES.

XX Fleischmann RD, White OR, Fraser CM, Venter JC;

XX WPI; 2001-647261/74.

XX Evaluating strain variation of Mycobacterium tuberculosis, comprises  
 PT determining the nucleotide sequence of the strain at positions in the  
 PT genome corresponding to positions where M. tuberculosis strains CDC  
 PT 1551 and H37Rv differ  
 XX  
 PS Claim 3; SEQ ID NO 1; 3pp + Sequence Listing; English.

XX The invention relates to evaluating strain variation within and between  
 CC different populations of the tuberculosis bacterial pathogen,  
 CC Mycobacterium tuberculosis or related Mycobacterium by determining the  
 CC nucleotide sequence of the first strain at positions in the complete  
 CC sequence of the genome that correspond to positions that differ in the  
 CC nucleotide sequences of M. tuberculosis strains CDC 1551 (AAI99683) and  
 CC H37Rv (AAI99682). The method is useful for evaluating strain variation of  
 CC M. tuberculosis and has valuable application in the fields of  
 CC tuberculosis genetics, epidemiology, patient treatment and epidemic  
 CC monitoring.

CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from USPTO  
 CC at seqdata.uspto.gov/sequence.html?docid=6294328B1.

XX SQ Sequence 4411529 BP; 758565 A; 1449983 C; 1444602 G; 758379 T; 0 other;

Query Match 12.0%; Score 173.8; DB 22; Length 4411529;  
 Best Local Similarity 51.7%; Pred. No. 1.4e-22;  
 Matches 499; Conservative 0; Mismatches 427; Indels 39; Gaps 3;

QY 273 GTACAGCCCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 332  
 DB 2416179 GCCAGTCCCGGCTTCTCGCGCAACCCCTACTGCGCGCGCGCGCGGGGTG 2416120

QY 333 CGCATCTCCGCTGACATCTCTTCCGCCG-----CGAGGCG 371

DB 2416119 CCGATCTGGGTGACGTGGAGTTAGCTGCGCGCTAGACGACGCGGCTCTACGGACCG 2416060

QY 372 AAGGCCCGATGCTGCGCATCCAGCGTTCACGCGAAGACCGCTGACCCCTGGTG 431

DB 2416059 CCGCGCAGCTGGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 2416000

QY 432 GCGAAATGCGGTGGCGCGGCAAGGCTGTGCGCGGCGCAACCTCGGCACCCG 491  
 DB 2415999 CAGCCATGCTGATCGCGCGGTGGCGCGCGCTGTGCGGCAATATCGGCAGTGG 2415940  
 QY 492 GCGCTCGACCTGCTGGCGGACGACATCGAGCTGTAGCTGTGGAGCTGTGCGAGCTTCCAG 551  
 DB 2415939 GTGCTGGATGTGCTGGAGCGCGCGAGCTGTGCGCGTGTGAGTGTGCTCCAGTTCCAG 2415880  
 QY 552 CTGGAAACCTGGATCGCTCAACCGCGAGGTGGCGCGGTGCTCAAGCTCAGCGAAGC 611  
 DB 2415879 CTGCACTGGCGCCCTGCTGCGCGCGCGAGCGCGGTGCTCAACATTGCGGAAGC 2415820  
 QY 612 CATATGATGCTAGGAGGCTGCTGACTACCACTGGCGCAACCGCGATCTTCGCG 671  
 DB 2415819 CACCTGGAGTGCATGCGACGATGCGCAATACACCGCGCGCGGTGTGAC 2415760  
 QY 672 GGTGCGCGGAGGTGCTGTGTAATCGCGCGGTGCTGACCGCGGTGATGCGCGGAT 731  
 DB 2415759 GCGCGGGTAGCGGTGGCGGGTGGATGACAGCGCGGCGCGCTGCTGCGACGCTCA 2415700  
 QY 732 ACCGTGCGGTGCTGCTGCG---TTGCGCTGACACAGCGCGCTTCAAGGCTTTCGCGCTG 788  
 DB 2415699 CCGCGCGAGTGGCGGTGCTGCTGCGCTGCGCGAGCGCGCGCGGGAATGCGCGTG 2415640  
 QY 789 ATCGAGGAAGCGCGCGAGAGTGGCTGCGCTTCCAGTTCGACAAGCTGTGCGGTGTCG 848  
 DB 2415639 CCGAGCGCCACTGCTGATGCTGCGCTTCCGAGACTGACGCTGCTGCGGTGCGG 2415580  
 QY 849 GAACCTGAAGATCGGTGGCGCGCACAACTATTCCACAGCGCTGCGCGCTGCGGTGCGG 908  
 DB 2415579 TCGATACCGGTGCCAGGTCCGCTGCGCTGCGCTGCGCGCTGCGCGCGGTGCGG 2415520  
 QY 909 CATGCGGTGCGCTGCGCTGCGCGCGCTGCGCGCGCTGCGCGCGGTGCGCGCGGTG 968  
 DB 2415519 CGCTCGGTGCGGTGCGCGCGGTGCGCGCGGTGCGCGCGGTGCGCGCGGTGCGG 2415460  
 QY 969 GCTCATGCTGCGCTGCGCGCGGTGCGCGCGGTGCGCGCGGTGCGCGCGGTGCGCGG 1028  
 DB 2415459 CGACACCGCGCGAGGTGCTGCGCGGTGCGCGCGGTGCGCGCGGTGCGCGCGGTG 2415400  
 QY 1029 GCGACACAGCTGCGCGCGGTGCGCGCGGTGCGCGCGGTGCGCGCGGTGCGCGCGGT 1088  
 DB 2415399 GCGACCAACCGCGCGCGGTGCGCGCGGTGCGCGCGGTGCGCGCGGTGCGCGG 2415355  
 QY 1089 AAGCTGTGCTGCTGCGCGCGGTGCGCGCGGTGCGCGCGGTGCGCGCGGTGCGCGG 1148  
 DB 2415354 AGGCTGTGATGATCGCGCGGTGCGCGCGGTGCGCGCGGTGCGCGCGGTGCGCGG 2415295  
 QY 1149 CCGGTGCGCGGTGCTGCGCGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1208  
 DB 2415294 GCGATGCGCTGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2415235  
 QY 1209 GAGC 1213  
 DB 2415234 GAGC 2415230  
 RESULT 12  
 ID AA199683/c  
 AC AA199683;  
 DT 15-JAN-2002 (first entry)  
 XX Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 2.  
 DE Mycobacterium tuberculosis; strain H37Rv; patient treatment; epidemic monitoring; ds.  
 KW Mycobacterium tuberculosis; strain H37Rv; patient treatment; epidemic monitoring; ds.  
 XX Mycobacterium tuberculosis.  
 OS Mycobacterium tuberculosis.

US294328-B1.  
 XX 25-SEP-2001.  
 XX 24-JUN-1998; 98US-0103840.  
 XX 24-JUN-1998; 98US-0103840.  
 XX (GENO-) INST GENOMIC RES.  
 PA Fleischmann RD, White OR, Fraser CM, Venter JC;  
 PI WPI; 2001-647261/74.  
 DR Evaluating strain variation of Mycobacterium tuberculosis, comprises  
 PT determining the nucleotide sequence of the strain at positions in the  
 PT genome corresponding to positions where M. tuberculosis strains CDC  
 PT 1551 and H37Rv differ  
 XX  
 PS Claim 4; SEQ ID NO 2; 3pp + Sequence Listing; English.  
 XX  
 CC The invention relates to evaluating strain variation within and between  
 CC different populations of the tuberculosis bacterial pathogen,  
 CC Mycobacterium tuberculosis or related Mycobacterium by determining the  
 CC nucleotide sequence of the first strain at positions in the complete  
 CC sequence of the genome that correspond to positions that differ in the  
 CC nucleotide sequences of M. tuberculosis strains CDC 1551 (AA199683) and  
 CC H37Rv (AA199682). The method is useful for evaluating strain variation of  
 CC M. tuberculosis and has valuable application in the fields of  
 CC tuberculosis genetics, epidemiology, patient treatment and epidemic  
 CC monitoring.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from USPTO  
 CC at seqdata.uspto.gov/sequence.html?DocID=6294328B1.  
 XX  
 SQ Sequence 4403765 BP; 757105 A; 1447799 C; 1441301 G; 757371 T; 189 other;  
 Query Match 11.9%; Score 172.2; DB 22; Length 4403765;  
 Best Local Similarity 51.6%; Pred. No. 2.7e-22;  
 Matches 498; Conservative 0; Mismatches 428; Indels 39; Gaps 3;  
 QY 273 GTACGCCCCGGGTGCTGCTGCGTACGACCCCTGCTGCTGCTGCTGCTGCTGCTGCTG 332  
 DB 2414838 GCGATCGCGGTGCTGCTGCGTACGACCCCTGCTGCTGCTGCTGCTGCTGCTG 2414779  
 QY 333 CGCATCTCGGTGACATGATCTCTTCGCCCG-----CGAGGCG 371  
 DB 2414778 CGCATCTCGGTGACATGATCTCTTCGCCCG-----CGAGGCG 371  
 QY 372 AAGGCCCCGATGCTGCGCATCATCCGTTTCCACGCGAAGAGCAGCAGTGTACCCCTGCTG 431  
 DB 2414718 CCGCGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2414659  
 QY 432 GCGCAATGCGGTGCGCGCGGACAAAGCTGTGCGCGTGGCGGCAACCTCGGCACCCCG 491  
 DB 2414658 CACGCGATGCTGATGCGCGGTGCGCGCGGCTGCTGCTGCGGCAATATCGCATGCG 2414599  
 QY 492 GCGCTCGACCTGCTGCGCGACGACATCGAGCTGTAGCTGTGCTGCTGCTGCTGCTGCTG 551  
 DB 2414598 GTGCTGGATGCTGCGACGAGCGCGCGAGCTGCTGCGGTGCTGCTGCTGCTGCTGCTG 2414539  
 QY 552 CTGGAAACCTGGATCGCTCAACCGCGAGGTGGCGGACCGTGTGCTGCTGCTGCTGCTGCTG 611  
 DB 2414538 CTGCACTGGCGCCCTGCTGCGCGCGCGAGCGCGGTGCTCAACATTGCGGAAGC 2414479  
 QY 612 CATATGATGCTAGGAGGCTGCTGACTACCACTGGCGCAACCGCGATCTTCGCG 671  
 DB 2414478 CACCTGGAGTGCATGCGACGATGCGCAATACACCGCGCGGTGCTGCTGCTGCTGCTG 2414419  
 QY 672 GGTGCGCGGAGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 731  
 DB 2414418 GCGCGGGTAGCGGTGCGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2414359



QY 958 TTTCCGGCTGCTCATCTGCCAGTGGGTACGGAGCGCGGCGTGGAGTACTACG 1017  
DB 911 TCCAAGGCTTGGCCACCGCTGTGGAATAATCGCGAGAAACCGCGTGGTATTATCG 970  
QY 1018 AGATTCCAAAGCCACCAACGTCGGCGCGCCCTGGCGCGATCGAGGGGCTGGGCGG 1077  
DB 971 ACACAGCAAAAGGCACGAATGTCGGCGCGACCGCGCGCGAT-----TGCGG 1018  
QY 1078 ACATCGACGGCAAGCTGTGCTCGCGCGCGAGACGGCAAGGCGCGGATTTCCATG 1137  
DB 1019 GTTTCGAACCCGCTCTCTCGATTTGGCGCGCATGGCCAAAGGAGGACTTCAAGC 1078  
QY 1138 ACTCGCGAGCGGTCGCGCGCTTCTGTCGGCGCGTGTACTGTGTCGCGTACGCGG 1197  
DB 1079 CCTCGCGACGCGTTGAAGATAAGGCAAAAGCGTGTCTCTGATCGCGTCGATCGC 1138  
QY 1198 GCGTATTGCCAGCAGCTG---GGCAAGCGGTCACCGTGTGCGCGTCGCAAGCTGG 1254  
DB 1139 CGCAATCGCGCGGATTTGGACGCGTGTGAACCTTGACCGACTGCGTCACTTTGG 1198  
QY 1255 ACGAAGCAGTCGCGCAGCGCGCGAGCTGGCGCGGAGACGGCGCGGATTTGTCGC 1314  
DB 1199 AGAGCGGTTTACAGCGGATACGCCCAAGCGGCGGATATTGCTTGTCTACGCC 1258  
QY 1315 CGCGCTGCGCGAGCCTGGACATGTTCAAGAACTTCGAAGAACGCGGACGCGCTGTCGCCA 1374  
DB 1259 CGCGCTGCGCGAGTTTCGATATGTTTAAAGGCTACGCGCACCGGTTTCGGAAGTGTATCG 1318  
QY 1375 AAGCGTAGAGG 1386  
DB 1319 AAGCGTTAAGG 1330

RESULT 14

AAS89816  
ID AAS89816 standard; cDNA; 4942 BP.

XX AAS89816;

XX 13-FEB-2002 (first entry)

XX DNA encoding novel human diagnostic protein #25620.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX Homo sapiens.

XX WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

XX 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

XX P-PSDB; ABG25629.

XX New isolated polynucleotide and encoded polypeptides, useful in  
XX diagnostics, forensics, gene mapping, identification of mutations  
XX responsible for genetic disorders or other traits and to assess  
XX biodiversity -

XX Claim 1; SEQ ID NO 25620; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and

XX polypeptide (II) sequences. (I) is useful as hybridisation probes,

CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags.  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: the sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX

XX Sequence 4942 BP; 1109 A; 1296 C; 1387 G; 1150 T; 0 other;

Query Match 10.9%; Score 158; DB 23; Length 4942;

Best Local Similarity 57.1%; Pred. No. 1e-19;

Matches 312; Conservative 0; Mismatches 225; Indels 9; Gaps 1;

QY 80 CATCGTTGTCGCGCTCGCAAGCGCGCATGTCCCTGTGTGCGTACTACGCGCGCGG 139

DB 1749 CGTCATATTGCGCTGGGCTTACCAGGCTTCTTCGTGTGGACTTTTCTCGCTCGCGG 1808

QY 140 CTTGCTTTTCGCGCTGTGATACCCGAGAGAACCCGCGGAGCTGGCCACCCTCGGTGC 199

DB 1809 TGTGACGCGCGCTTATGATACGATACACCGCTGCGCTGTGATAAATACCGC- 1867

QY 200 CCAGTATCCGAGGTGGAAGTGGTTCGCGGGAACCTCGAGCGGAGTCTCTCTGCTCCGC 259

DB 1868 -----AAGCCGTAGAACGCCACACGCGGCGTGTGAATGATGCTGATGCGCGC 1919

QY 260 CGCGAACTCTATGTACGCGCGCGCTTGTGCTGCGCACCCCTGCGCTGTACAGGCGCG 319

DB 1920 AGATCTGATTGTCGCCAGTCCCGGTATTGCTGCGCATCCATCCTTAAGCGCTGCGCG 1979

QY 320 CGCGAAAGCGTGGCGATCTCCGGTGACATCGATCTCTTCGCGCGGAGCGGAAAGCGCC 379

DB 1980 TGATGCGGGAATCGAAATCGTTGGCGATATCGAGCTGTCTGTGCGGAAAGCAAGCACC 2039

QY 380 GATCGTCGCCATCACCGGTTCCACGCGAAGACCGCTGACCCACCCTGCTGGGCGAAT 439

DB 2040 GATTGTGGCGATTACCGGTTCTAAGGGCAAAAGCAGGTCACCCAGCTAGTGGGTGAAT 2099

QY 440 GCGGTGGCGCGGACAAGCGTGTGCGCGCGGCAACCTCGGCGACCCCGCGCGCTCGA 499

DB 2100 GCGAAAGCGCGGGGTTAACGTTGGTGTGGTGCATATTGGCTGCTGCTGCTGAT 2159

QY 500 CTTGCTGGCGCGACGATCGAGCTGTACGTTGAGCTGTGAGCTGTCCAGCTGGAAC 559

DB 2160 GCTACTGGATGATGAGTGTGAACCTGTACGTTGGAACCTGTGAGCTTCCAGCTGGAAC 2219

QY 560 CTGCGATCGCTCAACGCGGAGTGGCGACCGTGTGACCTGACGAGACCATATGA 619

DB 2220 CACCTCCAGCTTACAGGCGGTAGACGCGACCATTTCTGAACGCTGACTGATATGA 2279

QY 620 TCGCTA 625

DB 2280 TCGCTA 2285

RESULT 15

AAS89273

ID AAS89273 standard; cDNA; 4944 BP.

XX AAS89273;

XX AC AAS89273;

XX



DT 13-FEB-2002 (first entry)  
 DE DNA encoding novel human diagnostic protein #25077.  
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
 XX Homo sapiens.  
 OS WO200175067-A2.  
 PN 11-OCT-2001.  
 PD 30-MAR-2001; 2001WO-US08631.  
 XX 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 XX (HYSE-) HYSEQ INC.  
 PA Drmanac RT, Liu C, Tang YT;  
 XX P-PSDB; ABG25086.  
 DR WPI; 2001-639362/73.  
 XX P-PSDB; ABG25086.  
 XX New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 XX Claim 1; SEQ ID No 25077; 103pp; English.  
 PS The invention relates to isolated polynucleotide (I) and  
 XX polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. AAS64197-AAS94564 represent novel human  
 CC diagnostic coding sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX Sequence 4944 BP; 1110 A; 1302 C; 1383 G; 1149 T; 0 other;  
 SQ

Query Match 10.9%; Score 158; DB 23; Length 4944;  
 Best Local Similarity 57.1%; Pred. No. 1e-19;  
 Matches 312; Conservative 0; Mismatches 225; Indels 9; Gaps 1;

QY 80 CATCGTTGTGGCGCCGAGAGCGCATCTCCCTGGTGGCTACCTGGCGCCCGCGG-139  
 DB 1749 CGTCATTATCGCCCTGGCGCTACCGGGCTTCTCGTGGCTACCTGGCTGGCGG 1808  
 QY 140 CTTGCCCTTTCCCGCTGGTTCGATACCCGAGAGAACCCCGGAGCTGGCCACCCCTGCGTGC 199  
 DB 1809 TGTGACGCCGCGCTTATGGATACGCGTATGACACCGCTGGCTGGATTAATACCG- 1867  
 QY 200 CCAGTATCCGAGTGGAGTGGCTTGGGGGAACTCGACCGCGAGTTCCTCTGCTCCGC 259  
 DB 1868 -----AAGCCGTAGAACGCCACACGGGCGAGTCTGAATGATGAATGGCTGATGGCGG 1919  
 QY 260 CCGGGAACCTCTATGTCAGCCCGCGCTGTGCTCGCGCACCCCTGCGGTGACAGGCCGC 319

DB 1920 AGATCTGATTGTCGCCAGTCCCGGTATTGCACTGGCGCATCCATCTTAAGCGCTGCCGC 1979  
 QY 320 CGCGAAAGGCGTGGCGCATCTCCGGTGACATCTCTTCGCCCGCGAGGCGAAGGCCCC 379  
 DB 1980 TGATGCCGGAATCGAATTCGTTGGGATATCGAGCTGTTCGTCCGGAAGCACAAGCACC 2039  
 QY 380 GATCGTCCCATCATCACCGGTTCCAAACGCGAAGAGCACCGTGACCACTGGTGGGCGAAAT 439  
 DB 2040 GATTGTGGCGATTACCGGTTCTAAGGGCAAAAGCACGCTACCCAGCTAGTGGGTGAAT 2099  
 QY 440 GCGGTTGGCCCGGAGACAGCGTGTCCGCGTGGGGGCAACCTCGGCACCCCGCGCTCGA 499  
 DB 2100 GCGAAAGCGCGCGGGGTTAACTGTTGGTGGTGGCAATATTGGCCTGCTGCGCTTGTAT 2159  
 QY 500 CCTGCTGGCCGACGACATCGAGCTGTAGCTGTGGAGCTGTGAGCTTCCAGCTGGAAC 559  
 DB 2160 GCTACTGGATGATGAGTGTGAACGTGTAGCTGTGAACGTGTGAGCTTCCAGCTGGAAC 2219  
 QY 560 CTGCGATCGCCTCAACCGCGAGGTGGCGACCGTGTGAACGTGACGAGAACCACTATATGA 619  
 DB 2220 CACCTCCAGCTTACAGGCGGTAGCAGCGACCATTTCTGAACGTGACTGAAGATCATATGA 2279  
 QY 620 TCGCTA 625  
 DB 2280 TCGCTA 2285

Search completed: August 14, 2003, 01:55:30  
 Job time : 443 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 14, 2003, 07:06:25 ; Search time 92 seconds

(without alignments)  
2149.344 Million cell updates/sec

Title: US-09-701-229-2

Perfect score: 2260

Sequence: 1 MSIIASHDFRIVVGLKSGM.....MFKNFERGRFLAKAVEELA 448

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 712578

Minimum DB seq length: 25

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=framet\_p2n.model -DEV=xlp  
-O/cgn2\_1/USPTO.spool/US09701229/runat\_09082003.161158.4405/app\_query.fasta.1.647  
-DB-Issued\_Patents\_NA\_QFMT-fastop -SUFFIX-p2n.rni -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS-human40.cdi  
-LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=25 -MAXLEN=2000000000  
-USER=US09701229.ecgn\_1.1.85 -runat\_09082003.161158.4405 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued\_Patents\_NA.\*  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/6C\_COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2260	100.0	1371	4	US-09-252-991A-7702
c 2	2252	99.6	1401	4	Sequence 7702, Ap
3	1148.5	50.8	1416	4	Sequence 7861, Ap
c 4	951	42.1	567	4	Sequence 1282, Ap
5	879	38.9	1830121	4	Sequence 7787, Ap
6	879	38.9	1830121	4	Sequence 1, Appli
c 7	626	27.7	564	4	US-09-557-884-1
8	603.5	26.7	2193	2	US-09-643-990A-1
9	603.5	26.7	2193	2	Sequence 7928, Ap
c 10	594.5	26.3	6693	4	US-08-934-481-1
11	591	26.2	360	4	Sequence 1, Appli
12	586.5	26.0	1350	1	US-09-290-602-1
					Sequence 195, App
					Sequence 7552, Ap
					Sequence 1, Appli

13	586.5	26.0	1350	1	US-08-665-435A-3	Sequence 3, Appli
14	586.5	26.0	1350	2	US-08-843-309-3	Sequence 3, Appli
15	586.5	26.0	1353	2	US-08-843-309-1	Sequence 1, Appli
c 16	525	23.2	4411529	3	US-09-103-840A-1	Sequence 2, Appli
c 17	515	22.8	4403765	3	US-09-103-840A-2	Sequence 1574, Ap
18	504.5	22.3	1368	4	US-09-107-532A-1574	Sequence 109, App
c 19	493	21.8	1003	3	US-08-961-083-109	Sequence 109, App
20	493	21.8	1003	3	US-09-536-784-109	Sequence 1, Appli
21	470	20.8	1372	4	US-09-530-836-1	Sequence 1001, Ap
22	469	20.8	1374	4	US-09-134-001C-1001	Sequence 1, Appli
23	421	18.6	1242	3	US-09-147-928-1	Sequence 1, Appli
24	395	17.5	2423	3	US-08-714-918-86	Sequence 86, Appli
25	395	17.5	2423	3	US-09-265-315-86	Sequence 86, Appli
26	395	17.5	2423	3	US-09-265-315-86	Sequence 86, Appli
27	395	17.5	2423	3	US-09-266-417-86	Sequence 86, Appli
28	391.5	17.3	1230025	4	US-09-198-452A-1	Sequence 1, Appli
29	389.5	17.2	2402	4	US-09-221-017B-881	Sequence 881, App
30	379	16.8	801	2	US-08-934-481-3	Sequence 3, Appli
31	379	16.8	801	4	US-09-290-602-3	Sequence 3, Appli
32	321	14.2	1359	4	US-09-252-991A-7624	Sequence 7624, Ap
33	321	14.2	2393	3	US-09-230-380-8	Sequence 8, Appli
c 34	292	12.9	843	4	US-08-961-527-356	Sequence 356, App
35	240	10.6	1513	4	US-09-252-991A-7701	Sequence 7701, Ap
c 36	210.5	9.3	1215	4	US-09-252-991A-7862	Sequence 7862, Ap
37	207	9.2	1360	2	US-08-732-612-1	Sequence 1, Appli
38	195.5	8.7	1479	4	US-09-328-352-897	Sequence 897, App
39	195.5	8.7	1542	4	US-09-252-991A-7625	Sequence 7625, Ap
c 40	195.5	8.7	1701	4	US-09-252-991A-7927	Sequence 7927, Ap
41	195	8.6	1503	4	US-09-328-352-2140	Sequence 2140, Ap
42	182.5	8.1	1717	4	US-08-936-165A-78	Sequence 78, Appli
43	182.5	8.1	1452	4	US-09-328-352-1894	Sequence 1894, Ap
44	173	7.7	536	3	US-09-147-928-3	Sequence 3, Appli
45	170.5	7.5	1400	3	US-09-144-918-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1

US-09-252-991A-7702  
; Sequence 7702, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 7702  
; LENGTH: 1371  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-7702

Alignment Scores:  
Pred. No.: 3.5e-241 Length: 1371  
Score: 2260.00 Matches: 448  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 4 Gaps: 0

US-09-701-229-2 (1-448) x US-09-252-991A-7702 (1-1371)

Qy 1 MetSerLeuIleAlaSerAspHisPheArgIleValValcGlyLeuGlyLysSerGlyMet 20  
|||||  
Db 25 ATGAGCGTCGATCGCTCCGACCACTCCCGATCGTTGTTCGCGCAAGCGGCATG 84

```
QY 21 SerLeuValArgTyrLeuAlaArgArgGlyLeuProPheAlaValAlaAspThrArgGlu 40
DB 85 TCCTTGGTGGCTTACCTGGCGCGCGCGCTTTCCTTTTCGCCGTTGTCCATACCGAGAG 144
QY 41 AsnProProGluLeuAlaThrLeuArgAlaGlnTyrProGlnValGlnValArgCysGly 60
DB 145 AACCCGCGGAGCTGGCCACCTCGGTGCCAGATATCCGACAGTGAAGTGGTTCGGC 204
QY 61 GluLeuAspAlaGluPheLeuCysSerAlaArgGluLeuTyrValSerProGlyLeuSer 80
DB 205 GAACTCGAGCGGAGTTCCTCTGCTCCGCCCGCAACTCTACGTCAGCCCGGCTGTGCG 264
QY 81 LeuArgThrProAlaLeuValGlnAlaAlaLysGlyValArgIleSerGlyAspIle 100
DB 265 CTGCGCACCCCGCGCTGTATACGCGCGCGGAAAGTGTGCGCATCTCCGGTGACATC 324
QY 101 AspLeuPheAlaArgGluAlaLysAlaProIleValAlaIleThrGlySerAsnAlaLys 120
DB 325 GATCTCTTCGCCGCGGAGCGGAGCGCGCATGCTGCGCATCACCGTTCACCGCGAG 384
QY 121 SerThrValThrLeuValGlyGluMetAlaValAlaAlaAspLysArgValAlaVal 140
DB 385 AGCACCGTCACCACTGTGTGGCGGAAATGGCGGTGGCGCGCAAGCGTGTCCGCGTC 444
QY 141 GlyGlyAsnLeuGlyThrProAlaLeuAspLeuLeuAlaAspIleGluLeuTyrVal 160
DB 445 GCGCGCAACTCGGCACCCCGCGCTCGACCTGCTGGCGGACGATCGAGCTGTACGTG 504
QY 161 LeuGluLeuSerSerPheGlnLeuGluThrCysAspArgLeuAsnAlaGlnValAlaThr 180
DB 505 TTGGAGCTGTCAGCTTCAGCTGGAACCTCGCATCGCTCAACCGCAGGTGGCGACC 564
QY 181 ValLeuAsnValSerGluAspHisMetAspArgTyrAspGlyMetAlaAspTyrHisLeu 200
DB 565 GTGCTGAACGTCAGCGAAGACCATATGGATCGCTACGACGGCATGCTACCACTG 624
QY 201 AlaLysHisArgIlePheArgGlyAlaArgGlnValValAlaAsnArgAlaAspAlaLeu 220
DB 625 GCGAAGCACCGGATCTTCGCGGTGCGCGCCAGGTCGTGTGAATCGCGCGCATGCCCTG 684
QY 221 ThrArgProLeuIleAlaAspThrValProCysTrpSerPheGlyLeuAsnLysProAsp 240
DB 685 ACCCGACGCTGATCGCGATACCGTGGCGTGTGCTGCTGCTGCGGCTGAACAGCGGAC 744
QY 241 PheLysAlaPheGlyLeuIleGluAspGlyGlnLysTrpLeuAlaPheGlnPheAsp 260
DB 745 TTCAAGGCTTCGCGCTGATCGAGGAGACGCGCCAGAGTGGCTGGCGTTCAGTTCGAC 804
QY 261 LysLeuLeuProValGlyGluLeuLysIleArgGlyAlaHisAsnTyrSerAsnAlaLeu 280
DB 805 AAGCTGCTGCGCGGTTGGCGAATGAAGATCCGTCGCGCCCACTATTTCAACGCGCTC 864
QY 281 AlaAlaLeuAlaLeuGlyHisAlaValGlyLeuProPheAspAlaMetLeuGlyAlaLeu 300
DB 865 GCGCGCTGGCGTGGCCCACTGGCTGCGCTGCGCTGCGGCGCATGCTGCGCGCGCTG 924
QY 301 LysAlaPheSerGlyLeuAlaHisArgCysGlnTrpValArgGluArgGlnGlyValSer 320
DB 925 AAGCGGTTTCGCGCTGCTCATCGCTGCCAGTGGGTACGCGACGCGCGGCGGTGAGC 984
QY 321 TyrTyrAspAspSerLysAlaThrAsnValGlyAlaAlaLeuAlaAlaIleGluGlyLeu 340
DB 985 TACTACGACGATTTCAAGGCCACCAACGTCGCGCGCGCGCTGCGCGCGCATCGAGGGCTG 1044
QY 341 GlyAlaAspIleAspGlyLysLeuValLeuLeuAlaGlyLysGlyValAlaAsp 360
DB 1045 GGTGCGACATCGAGCGCAAGCTGTGTGCTGCGCGCGGAGACGCGCAAGGGCGCGCAT 1104
QY 361 PheHisAspLeuArgGluProValAlaAlaPheCysArgAlaValValLeuLeuGlyArg 380
DB 1105 TTCCATGACCTCGCGGACCGGTCGCGCGCTTCTCGCGCGGTGTACTGCTTGGCGGT 1164
QY 381 AspAlaGlyLeuIleAlaGlnAlaLeuGlyAsnAlaValProLeuValArgValAlaThr 400
```

```
DB 1165 GACGCGGCGCTGATTGCCAGGCACTGGGCAACGCGGTACCGCTGGTGGCGTGCACAG 1224
QY 401 LeuAspGluAlaValArgGlnAlaAlaGluLeuAlaArgGluGlyAspAlaValLeuLeu 420
DB 1225 CTGGACGAGAGCACTCCGCGAGCCCGCGAGCTGGCGCGGAGGCGATGCGGTGTGTTG 1284
QY 421 SerProAlaCysAlaSerLeuAspMetPheLysAsnPheGluGluArgGlyArgLeuPhe 440
DB 1285 TCGCGCGCGCTGGCGGAGCTGGACATGTTCAAGAACTTCGAAGAACGCGGACGCGCTGTC 1344
QY 441 AlaLysAlaValGluGluLeuAla 448
DB 1345 GCCAAAGCGGTAGAGGAGTAGCG 1368

RESULT 2
US-09-252-991A-7861/C
; Sequence 7861, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ IDS: 33142
; SEQ ID NO 7861
; LENGTH: 1401
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-7861

Alignment Scores:
Pred. No.: 2,8e-240 Length: 1401
Score: 2252.00 Matches: 446
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.65% Indels: 0
DB: 4 Gaps: 0

US-09-701-229-2 (1-448) x US-09-252-991A-7861 (1-1401)
QY 1 MetSerLeuIleAlaSerAspHisPheArgIleValValGlyLeuGlyLysSerGlyMet 20
DB 1338 ATGAGCCTGATCGGCTCCGACCACTTCGCGCATCGTTGCGGCCCTCGGCAAGCGGCAATG 1279
QY 21 SerLeuValArgTyrLeuAlaArgArgGlyLeuProPheAlaValAlaAspThrArgGlu 40
DB 1278 TCCCTGGTGGCTACCTGGCGCGCGCGCTTTCGCGCTTGTGATACCGGAG 1219
QY 41 AsnProProGluLeuAlaThrLeuArgAlaGlnTyrProGlnValGluValArgCysGly 60
DB 1218 AACCGCGGAGCTGGCCACCTCGTGCCTGCCAGTATCGCAGGTGGAAGTGGTTCGCGC 1159
QY 61 GluLeuAspAlaGluPheLeuCysSerAlaArgGluLeuTyrValSerProGlyLeuSer 80
DB 1158 GAATCGACGCGGAGTTCCTCTGCTCGCGCGCGCAACTCTACGTACGCGCGCGGTGTCG 1099
QY 81 LeuArgThrProAlaLeuValGlnAlaAlaLysGlyValArgIleSerGlyAspIle 100
DB 1098 CTGCGCACCCCGCGCTGGTACAGCGCGCGCGGAAAGGTGTGCGCATCTCCGGTGACATC 1039
QY 101 AspLeuPheAlaArgGluAlaLysAlaProIleValAlaIleThrGlySerAsnAlaLys 120
DB 1038 GATCTTCGCGCGGAGCGGCGGAGCGCGCATGCTCGCATACCGGTTCACGCGCAAG 979
QY 121 SerThrValThrLeuValGlyGluMetAlaValAlaAlaAspLysArgValAlaVal 140
```

```
Db 978 AGCACCGTGCACCCCTGGTGGCGAAATGCGGTGGCGGAGCAACGCGTGTCCGCGTC 919
Qy 141 GlyGlyAsnLeuGlyThrProAlaLeuAspLeuAlaAspAspIleGluLeuTyrVal 160
Db 918 GCGCGCAACCTCGGCACCCCGGCGCTGACCTGCTGGCGGACGACATCGAGCTGACGTG 859
Qy 161 LeuGluLeuSerSerPheGlnLeuGluThrCysAspArgLeuAsnAlaGluValAlaThr 180
Db 858 TTGGAGCTGTCCAGCTCCAGCTGGAACCTGCGCTCAACGCGGAGGTGGCGACC 799
Qy 181 ValLeuAsnValSerGluAspHisMetAspArgTyrAspGlyMetAlaAspTyrHisLeu 200
Db 798 GTGTGACGTCACGCAAGACCATATGATGCTACGACGCGCATGGCTGACTACCACTG 739
Qy 201 AlalysHisArgIlePheArgGlyAlaArgGlnValValAsnArgAlaAspAlaLeu 220
Db 738 GCCAAGCACCGGATCTCCGGGTCGCCGCGAGGTGCTGTGAATCGCGCGCATGCCGTG 679
Qy 221 ThrArgProLeuIleAlaAspThrValProCysTrpSerPheGlyLeuAsnLysProAsp 240
Db 678 ACCGAGCGCTGATCGCGGATACCGTGGCTGCTGCTGGCTGGAACGCGGAC 619
Qy 241 PheLysAlaPheGlyLeuIleGluAspGlyGlnLysTrpLeuAlaPheGlnPheAsp 260
Db 618 TTCAAGGCTTTCGGCCCTGATCGAGGAAGACGGCCAGAGTGGCTGGCGTTCCAGTTCCGAC 559
Qy 261 LysLeuLeuProValGlyGluLeuLysIleArgGlyAlaHisAsnTyrSerAsnAlaLeu 280
Db 558 AAGCTGCTGCGGTTGGGCACTCAAGATCGTGGCGGCCCACTATTCACAGCGCTC 499
Qy 281 AlaAlaLeuAlaLeuGlyHisAlaValGlyLeuProPheAspAlaMetLeuGlyAlaLeu 300
Db 498 GCCGCGTGGCGCTGGGCGCATGCGGCTGGCTGCGGCGCATGCTCGCGCGCGTGC 439
Qy 301 LysAlaPheSerGlyLeuAlaHisArgCysGlnTrpValArgGluArgGlnGlyValSer 320
Db 438 AAGCGCTTTCGGCCCTGGCTATCGCTGCGCGTGGGTACCGAGCGCGCGGCGTGCAGC 379
Qy 321 TyrTyrAspSerLysAlaThrAsnValGlyAlaAlaLeuAlaIleGluGlyLeu 340
Db 378 TACTACGACGATTCCAAAGCCACCAACGTCGCGCGCCCTGCGCGGATCGAGGGCTG 319
Qy 341 GlyAlaAspIleAspGlyLysLeuValLeuLeuAlaGlyLysGlyLysGlyAlaAsp 360
Db 318 GGTCCGACATCGACGCAAGCTGGTGTCTGCGCGCGGAGACGCGCAAGGCGCGCAT 259
Qy 361 PheHisAspLeuArgGluProValAlaArgPheCysArgAlaValValLeuGlyValArg 380
Db 258 TTCCATGACCTGCGGAGCGCGTCTGCGCGCTTCTGCGGCGGCGGTACTGCTGGCGGT 199
Qy 381 AspAlaGlyLeuIleAlaGlnAlaLeuGlyAsnAlaValProLeuValArgValAlaThr 400
Db 198 GACGCGGCGCTGATTCCCGACGCACTGGGCAACGCGGTACCGCTGGTGGCGGTGCGAACG 139
Qy 401 LeuAspGluAlaValArgGlnAlaAlaGluLeuAlaArgGluGlyAspAlaValLeuLeu 420
Db 138 CTGGACGAAGCAGTCCCGCAGCGCGCGAGTGGCCCGCAAGCGATGCGGCTGCTGTTG 79
Qy 421 SerProAlaCysAlaSerLeuAspMetPheLysAsnPheGluGluArgGlyArgLeuPhe 440
Db 78 TCGCGCGCTCGCGAGCGCTGGAGATGTTCAAGAACTTCGAAGAACCGCGGACGCGCTGTC 19
Qy 441 AlalysAlaValGluGlu 446
Db 18 GCCAAAGCGGTAGAGGAG 1
```

## RESULT 3

US-09-328-352-1282  
; Sequence 1282, Application US/09328352  
; Patent No. 6562958  
; GENERAL INFORMATION:  
; APPLICANT: Gary L. Breton et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER

; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: GTC99-03PA

; CURRENT APPLICATION NUMBER: US/09/328,352

; NUMBER OF SEQ ID NOS: 8252

; SEQ ID NO 1282

; LENGTH: 1416

; TYPE: DNA

; ORGANISM: Acinetobacter baumannii

US-09-328-352-1282

Alignment Scores: 8.45e-118 Length: 1416  
Pred. No.: 1148.50 Matches: 240  
Score: 1148.50  
Percent Similarity: 66.44% Conservative: 59  
Best Local Similarity: 53.33% Mismatches: 142  
Query Match: 50.82% Indels: 9  
DB: 4 Gaps: 4

US-09-701-229-2 (1-448) x US-09-328-352-1282 (1-1416)

```
Qy 3 LeuIleAlaSerAspHisPheArgIleValValGlyLeuGlyLysSerGlyMetSerLeu 22
Db 73 TTAATACACGTGGTGGATTAAAGATTGTAGCAGGCTTGGGAATATCAGGTGTTCTGCT 132
Qy 23 ValArgTyrLeuAlaArgArgGlyLeuProPheAlaValValAspThrArgGluAsnPro 42
Db 133 GTAATTTCTTCGTCATGAACAGGCTACCAAGTCTGCTGAACGGATTCCCGCCACACCT 192
Qy 43 ProGluLeuAlaThrLeuArgAlaGlnTyrProGln---ValGluValArgCysGlyGlu 61
Db 193 CCCGGA-----CAGCATCAGATTCCGGCTGGTGTAAACACGAGTTTGGTCAG 240
Qy 62 LeuAspAlaGluPheLeuCysSerAlaArgGluLeuTyrValSerProGlyLeuSerLeu 81
Db 241 CTGATCAAGAAATATTATTACAGCAGACAGAGATTATTAAAGCCAGGCTTGCACCA 300
Qy 82 ArgThrProAlaLeuValGlnAlaAlaLysGlyValArgIleSerGlyAspIleAsp 101
Db 301 CAATTACCGAAATTCAGGCAGCTATTGCTAAAGTATTCTGTGGTGGCGATATCCAG 360
Qy 102 LeuPheAlaArgGluAlaLysAlaProIleValAlaIleThrGlySerAsnAlaLysSer 121
Db 361 TTATTCGCGCTGCCACTGATGTCGCGATTGTCGATTCAGAGTTCCAAATGCAAAAAGT 420
Qy 122 ThrValThrThrLeuValGlyGluMetAlaValAlaAlaAspLysArgValAlaValGly 141
Db 421 ACTGTACCACTTTAATTTGGCTTAATGGCTAAGGATGCGAGTAAGAAAGTTGCGATAGC 480
Qy 142 GlyAsnLeuGlyThrProAlaLeuAspLeuAlaAspAspIleGluLeuTyrValLeu 161
Db 481 GGCAACCTTGGTCGACGACGCTTAGATTTACTTAAAGATCAACACGAGATTACTGGCTCT 540
Qy 162 GluLeuSerSerPheGlnLeuGluThrCysAspArgLeuAsnAlaGluValAlaThrVal 181
Db 541 GAGTTATCAAGCTTTCAGTTAGATTAGAACCCACCTCTCCTTAAACCGTGGAGTGGTGT 600
Qy 182 LeuAsnValSerGluAspHisMetAspArgTyrAspGlyMetAlaAspTyrHisLeuAla 201
Db 601 CTCATATGAGTGAAGACCATTTAGCCGTCTGGAATATGCTGGGGTATCATCAAGCA 660
Qy 202 LysHisArgIlePheArgGlyAlaArgGlnValValValAsnArgAlaAspAlaLeuThr 221
Db 661 AACATCGTATTTCCAAAGCGCTTAAAGAAAGTTGTTATTAAACGAGATGATCGGTAAAG 720
Qy 222 ArgProLeuIleAlaAspThrValProCysTrpSerPheGlyLeuAsnLysProAspPhe 241
Db 721 CGTCCACTTGTTCAGATACGACCAATGCAAGCTTGTGTTTAAACGACCGCGGATTTA 780
Qy 242 LysAlaPheGlyLeuIleGluGlu---AspGlyGlnLysTrpLeuAlaPheGlnPheAsp 260
Db 781 AATCAATATGCGCGTTTAAAGAGATGCCAGCGTACGCTTTGGCTTGTGCTGTTTACAG 840
```

QY 261 LysLeuLeuProValGlyGluLeuLysIleArgGlyAlaHisAsnTyrSerAsnAlaLeu 280  
Db 841 CGGTTAAATAAAGCTCAGATTTATATATTAATCAAGATGATGACCAATAGTAAATGCTTTA 900  
QY 281 AlaAlaLeuAlaLeuGlyHisAlaValGlyLeuProPheAspAlaMetLeuGlyAlaLeu 300  
Db 901 GCTTGTAGCATGGGGGAAGCAATGGTGTACCTATGAGTCAATGCTTGAACATTTA 960  
QY 301 LysAlaPheSerGlyLeuAlaHisArgCysGlnTrpValArgGluArgGlnGlyValSer 320  
Db 961 AAAAAATTTAAAGCATAGACACCGCTGTGAGTACGTTAAACCCGTGATGTCCT 1020  
QY 321 TyrTyrAspAspSerLysAlaThrAsnValGlyAlaAlaLeuAlaAlaIleGluGlyLeu 340  
Db 1021 TATTAAATGACTCTAAAGTACTAATGTTGGTCTACACTTGCACCAATTTGATGCTTA 1080  
QY 341 GlyAlaAspIleAsp-----GlyLysLeuValLeuAlaGlyGlyAspGlyLys 357  
Db 1081 GTGCTGCCATTAAGTGAAGTAAGGTGACCTTATTTAGTGGACAGGCCAA 1140  
QY 358 GlyAlaAspPheHisLeuArgGluProValAlaAlaArgPheCysArgAlaValValLeu 377  
Db 1141 GGACAAGATTTTCTCTTACGTCTTCTATCGAAATATATCCCAAGTTGTGCTATTG 1200  
QY 378 LeuGlyArgAspAlaGlyLeuIleAlaGlnAlaLeuGlyAsnAlaValProLeuValArg 397  
Db 1201 ATGGTGAAGATCGCGCTGTCATCGAACAGCCATTCAGAGCGCACTAAATTTTACAT 1260  
QY 398 ValAlaThrLeuAspGluAlaValArgGlnAlaAlaGluLeuAlaArgGluGlyAspAla 417  
Db 1261 GCAGCAACGCTTAAAGACGCTGCGAGCTGTGTCACGCTGAACACCAACGCTGAAGATGTG 1320  
QY 418 ValLeuLeuSerProAlaCysAlaSerLeuAspMetPheLysAsnPheGluGluArgGly 437  
Db 1321 GTATTGCTATCACCAGCATGTGCAAGCTTTGATATGTTTAAAGTTTATATGACCGTGT 1380  
QY 438 ArgLeuPheAlaLysAlaValGluGluLeu 447  
Db 1381 CAGCAGTTTGTGCTGCGTCAATTCGTTG 1410

RESULT 4  
US-09-252-991A-7787/c  
; Sequence 7787, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 7787  
; LENGTH: 567  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-7787  
  
Alignment Scores:  
Pred. No.: 1,72e-96 Length: 567  
Score: 951.00 Matches: 188  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 42.08% Indels: 0  
DB: 4 Gaps: 0  
  
US-09-701-229-2 (1-448) x US-09-252-991A-7787 (1-567)  
QY 242 LysAlaPheGlyLeuIleGluGluAspGlyGlnLysTrpLeuAlaPheGlnPheAspLys 261

Db 565 AAGGCTTTGGCGCTGATCGAGGAAGACGGCCAGAGTGGCTGGCTTCCAGTTCGACAAG 506  
QY 262 LeuLeuProValGlyGluLeuLysIleArgGlyAlaHisAsnTyrSerAsnAlaLeuAla 281  
Db 505 CTGCTCCCGGTGGCCAACTGAAGATCCGTGGCGGCCACAACTATTTCACACGGCTCGCC 446  
QY 282 AlaLeuAlaLeuGlyHisAlaValGlyLeuProPheAspAlaMetLeuGlyAlaLeuLys 301  
Db 445 GCGCTGGCGCTGGCCATCGGCTCGGCTGCCCTCGAGCCCATGCTCGGCGGCTGAAG 386  
QY 302 AlaPheSerGlyLeuAlaHisArgCysGlnTrpValArgGluArgGlnGlyValSerTyr 321  
Db 385 GCGTTTCCGCGCTGCTCATCGCTGCCAGTGGGTACGGAGCGGCGGCGGCTGAGCTAC 326  
QY 322 TyrAspAspSerLysAlaThrAsnValGlyAlaAlaLeuAlaAlaIleGluGlyLeuGly 341  
Db 325 TACGAGATTCCAAGGCCACCAACGTCGGCGCGCGCTGGCGCGCATCGAGGGGCTGGGT 266  
QY 342 AlaAspIleAspGlyLysLeuValLeuAlaGlyGlyAspGlyLysGlyAlaAspPhe 361  
Db 265 GCCGACATCGACGGCAAGCTGTGTCTCGCGCGGAGAGCGCAAGGCGCGCATTC 206  
QY 362 HisAspLeuArgGluProValAlaAlaArgPheCysArgAlaValValLeuLeuGlyArgAsp 381  
Db 205 CATGACCTCGCGAGCGGCTCGCGGCTTCTCGCGGCGGCTGCTGCTGCTGCTGCTGCTG 146  
QY 382 AlaGlyLeuIleAlaGlnAlaLeuGlyAsnAlaValProLeuValArgValAlaThrLeu 401  
Db 145 GCGGCGCTGTTGCCAGCAGCTGGCAACGCGGTACCGCTGGTGGCGCTGCCAACGCTG 86  
QY 402 AspGluAlaValArgGlnAlaAlaGluLeuAlaArgGluGlyAspAlaValLeuLeuSer 421  
Db 85 GACGAAGCAGTCCGCGAGCGCGCGAGCTGGCGCGGCGGCGGCGGCGGCGGCGGCGG 26  
QY 422 ProAlaCysAlaSerLeuAspMet 429  
Db 25 CCGGCTGCGCGAGCTGGACATG 2  
  
RESULT 5  
US-09-557-884-1  
; Sequence 1, Application US/09557884  
; Patent No. 6506581  
; GENERAL INFORMATION:  
; APPLICANT: Fleischmann et al.  
; TITLE OF INVENTION: The Nucleotide sequence of  
; the Haemophilus influenzae Rd Genome, Fragments  
; Thereof, and Uses Thereof  
; NUMBER OF SEQUENCES: 1  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: MD  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3 1/2 inch diskette  
; COMPUTER: Dell Pentium  
; OPERATING SYSTEM: MS DOS v6.22  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/557,884  
; FILING DATE: 25-Apr-2000  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/476,102  
; FILING DATE: JUN-5-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Michelle S. Marks  
; REGISTRATION NUMBER: 41,971  
; REFERENCE/DOCKET NUMBER: PB186P3  
; TELECOMMUNICATION INFORMATION:

TELEPHONE: 301-309-8504

TELEFAX: 301-309-8439

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1830121 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 1:

US-09-557-884-1

Alignment Scores:

Pred. No.: 4, 12e-83 Length: 1830121

Score: 879.00 Matches: 206

Percent Similarity: 60.72% Conservative: 63

Best Local Similarity: 46.50% Mismatches: 136

Query Match: 38.89% Indels: 38

DB: 4 Gaps: 10

US-09-701-229-2 (1-448) x US-09-557-884-1 (1-1830121)

Qy 12 ValValGlyLeuGlyLysSerGlyMetSerLeuValArgTyrLeuAlaArgGlyLeu 31  
 Db 1203162 ATCATCGGCTGGCAAAACAGGCTCTTCTGTGTGATATCTCTTATCCACAGGCT 1203221  
 Qy 32 ProPheAlaValAspThrArgGluAsnProProGluLeuAlaThrLeuArgAlaGln 51  
 Db 1203222 AATATCGTGTGATGATACGCAAAATCCTACTGCTATGATAAATCTT----- 1203272  
 Qy 52 TyrProGln---ValGluValArgCysGlyGluLeuAspAlaGluPheLeuCysSerAla 70  
 Db 1203273 ---CCTCAAAATATCCCTCTCATACGCTGGTGTGATTTAAATCAGGAATGCTTACTTTGAAAGC 1203329  
 Qy 71 ArgGluLeuTyrValSerProGlyLeuSerLeuArgThrProAlaLeuValGlnAlaAla 90  
 Db 1203330 GATATGATGTATATAGCCAGGCTTGGGGTAAAAACACCGAATTCAAACCGCACTT 1203389  
 Qy 91 AlalysGlyValArgIleSerGlyAspIleAspLeuPheAlaArgGluAlaAlaPro 110  
 Db 1203390 AAAGCGGAGTGAAGTAATCGCGGATATTGAATTTCTGCCGCGCAGCAACAGCCA 1203449  
 Qy 111 IleValAlaIleThrGlySerAsnAlaLysSerThrValThrLeuValGlyGluMet 130  
 Db 1203450 ATTTGGGGATTTACAGGTTCAATGGTAAAGTACCCGTAACCTACTTTAGTTTAAAGATG 1203509  
 Qy 131 AlaValAlaAlaAspLysArgValAlaValGlyGlyAsnLeuGlyThrProAlaLeuAsp 150  
 Db 1203510 CGGAAGCTGCTGTGTGAAGTTGGTATGGCGGAATATTTGGATTCCTCGCTTTGTCA 1203569  
 Qy 151 LeuLeuAlaAspIleGluLeuTyrValLeuGluLeuSerSerPheGlnLeuGluThr 170  
 Db 1203570 TTGTTGAATGAAGATTGTGAATTTATGTACTAGAGCTTCTACTTTTCAGCTTGAGACA 1203629  
 Qy 171 CysAspArgLeuAsnAlaGluValAlaThrValLeuAsnValSerGluAspHisMetAsp 190  
 Db 1203630 ACTTATAGCTTTAAAGCTGCGCAGCGACTGCTTGAACGCTGACTGAAGATCATATGGAT 1203689  
 Qy 191 ArgTyrAspGlyMetAlaAspTyrHisLeuAlaLysHisArgIlePheArgGlyAlaArg 210  
 Db 1203690 CGCTATATGATTTAGAGATTTATCGCCAGCAAAATACGCAATTTATCATATGCTTAAA 1203749  
 Qy 211 GlnValValValAsnArgAlaAspAlaLeuThr-----ArgProLeuIleAla 226  
 Db 1203750 GTAGTGTGTGTGAACATGAATAGATAGCTGCTTTTGGGAAACGAAATCAAGCGCAA 1203809  
 Qy 227 AspThrValProCysTrpSerPheGlyLeuAsnLysProAspPheLysAlaPheGlyLeu 246  
 Db 1203810 CATACCGCTT-----TCTTTTGGCGAAATAGTGGGAT-----TATTGGCTA 1203851  
 Qy 247 IleGluGluAspGlyGlnLysTrpLeuAlaPheGlnPheAspLysLeuLeuProValGly 266  
 Db 1203852 AAAACTGAATGGCAAGCAATATTTAATGTTAAAGATGAAGTGTATTTACCTTTGTGAA 1203911

Qy 267 GluLeuLysIleArgGlyAlaHisAsnTyrSerAsnAlaLeuAlaLeuAlaLeuGly 286  
 Db 1203912 GAAGCTACATTTGGTGTCTGCCATAATATATGAACATTTTGGCAGCAACAGCATTTGGCA 1203971  
 Qy 287 HisAlaValGlyLeuProPheAspAlaMetLeuLeuGlyAlaLeuLysAlaPheSerGlyLeu 306  
 Db 1203972 CAAGCTATAGTATTAAATTTAGATTCAATTCGTACCCGCACTTCGTCTCAATTCAAAGGGTTA 1204031  
 Qy 307 AlaHisArgCysGlnTrpValArgGlnGlyValSerTyrTyrAspAspSerLys 326  
 Db 1204032 GATCATCGTTTTCATTTAGTCATCAAGCTAATGGCATCTCTGGATTAATGACTCTAAA 1204091  
 Qy 327 AlaThrAsnValGlyAlaAlaLeuAlaAlaIleGluGlyLeuGlyAlaAspIleAspGly 346  
 Db 1204092 GCAACAAATGTGGGAGTACAGATTCGTCTGCTGGCTT-----TATATTGAGGGT 1204145  
 Qy 347 LysLeuValLeuLeuAlaGlyAspGlyLysGlyAlaAspPheHisAspLeuArgGlu 366  
 Db 1204146 AAATTGCATTTGTTGTAGCGGAGACGGAAGGGCTGATTTTTCAGAAATTAGCTGAA 1204205  
 Qy 367 ProValAlaArg-----PheCysArgAlaValValLeuLeuLeuGlyArgAsp 381  
 Db 1204206 TTAATTAATCAACACACATATTATTGTATTGT-----TTTGGTGCAGAT 1204250  
 Qy 382 AlaGlyLeuLeuAlaGlnAlaLeuGlyAsnAlaValProLeuValArgValAlaThrLeu 401  
 Db 1204251 GGTGCGTGTGTCAAAATTTTCATCGCAAGTTAT-----TTG 1204289  
 Qy 402 AspGluAlaValArgGlnAlaAlaGluLeuAlaArg-----GluGlyAspAla 417  
 Db 1204290 TTCGATCAATGGAACAGCGGATAGTAATTTTACGCCCAACATTTGCAAGCGGAGATATG 1204349  
 Qy 418 ValLeuLeuSerProAlaCysAlaSerLeuAspMetPheLysAsnPheGluGluArgGly 437  
 Db 1204350 GTATTATGTGCGCTGCTGTGCAAGTCTCGATCAGTTGCTTCTTTTGAAGCGCGC 1204409  
 Qy 438 ArgLeuPhe 440  
 Db 1204410 GAAGAATTT 1204418

## RESULT 6

US-09-643-990A-1

Sequence 1, Application US/09643990A

Patent No. 6528289

GENERAL INFORMATION:

APPLICANT: Robert D. Fleischmann

Mark D. Adams

Owen White

Hamilton O. Smith

J. Craig Venter

TITLE OF INVENTION: The Nucleotide sequence of

the Haemophilus influenzae Rd Genome, Fragments  
Thereof, and Uses Thereof

NUMBER OF SEQUENCES: 1

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville,

STATE: MD

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: 3 1/2 inch diskette

COMPUTER: Dell Pentium

OPERATING SYSTEM: MS DOS v6.22

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/643,990A

FILING DATE: 23-Aug-2000

CLASSIFICATION: &lt;Unknown&gt;

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/487,429

FILING DATE: 1995-06-07

APPLICATION NUMBER: 08/426,787  
 FILING DATE: 1995-04-21  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Kenley K. Hoover  
 REGISTRATION NUMBER: 40,302  
 REFERENCE/DOCKET NUMBER: PB186PIC1  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 301-610-5790  
 TELEFAX: 310-309-8439  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1830121 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
 US-09-643-990A-1

Alignment Scores:  
 Pred. No.: 4,12e-83 Length: 1830121  
 Score: 879.00 Matches: 206  
 Percent Similarity: 60.72% Conservativity: 63  
 Best Local Similarity: 46.50% Mismatches: 136  
 Query Match: 38.89% Indels: 38  
 DB: 4 Gaps: 10

US-09-701-229-2 (1-448) x US-09-643-990A-1 (1-1830121)  
 QY 12 ValValGlyLeuGlySerGlyMetSerLeuValArgTyrLeuAlaArgArgGlyLeu 31  
 DB 1203162 ATCATCGGCTGGCAACAGCTCTTCTGTGGGATTATCTTATCCCAACAGGCT 1203221  
 QY 32 ProPheAlaValValAspThrArgGluAsnProGluLeuAlaThrLeuArgAlaGln 51  
 DB 1203222 AATATTGCTGTGATGATCTGCAAAATCTACTGCTGATGATAAACTT----- 1203272  
 QY 52 TyrProGln---ValGluValArgCysGlyGluLeuAspAlaGluPheLeuCysserAla 70  
 DB 1203273 ---CCTCAAAATATCCCTCTCATCTGCTGATGTTAAATCAGGAATGGTACTTGAAGC 1203329  
 QY 71 ArgGluLeuTyrValSerProGlyLeuSerLeuArgThrProAlaLeuValGlnAlaLa 90  
 DB 1203330 GATATGATTGTTATTAGCCAGGCTTGGCGTAAACACCAAGAAATTCACACCCGACTT 1203389  
 QY 91 AlaLysGlyValArgIleSerGlyAspIleAspLeuPheAlaArgGluAlaLysAlaPro 110  
 DB 1203390 AAAGCGAGTGGGAAGTAAATCGCGATATTGAATTTCTGCGCGCGACACAAAGCCA 1203449  
 QY 111 IleValAlaIleThrGlySerAsnAlaLysSerThrValThrLeuValGlyGluMet 130  
 DB 1203450 ATTGTGGGATTACAGGTTCAATGTAAGTAAACGCTACCTGATTTATGAAATG 1203509  
 QY 131 AlaValAlaAlaAspLysArgValAlaValGlyGlyAsnLeuGlyThrProAlaLeuAsp 150  
 DB 1203510 GCGAAGCTGCTGCTGTAAGTGTGTGCGGCAATATTGGGATTCCTCCGCTTGCTCA 1203569  
 QY 151 LeuLeuAlaAspIleGluLeuTyrValLeuGluLeuSerSerPheGlnLeuGluThr 170  
 DB 1203570 TTGTTGAATGAAGATGTGAACCTTTATGTAAGAGCTTCTAGTTTTCAGCTTGAGACA 1203629  
 QY 171 CysAspArgLeuAlaGluValAlaThrValLeuAsnValSerGluAspHisMetAsp 190  
 DB 1203630 ACTTATAGCTTAAAGCTCGCGAGCTGCTTGAAGCTGACTGAAGATCATATGAT 1203689  
 QY 191 ArgTyrAspGlyMetAlaAspTyrHisLeuAlaLysHisArgIlePheArgGlyAlaArg 210  
 DB 1203690 CGCTATATGATTTAGAAATATATCCCAAGCAAAATACGCTATTTATCATATGCTAAA 1203749  
 QY 211 GlnValValValAsnArgAlaAspAlaLeuThr-----ArgProLeuIleAla 226  
 DB 1203750 GPAGTGTGTTGAACAATGAAGATAGGCTGACTTTTGGGCAAAACGAAATCAAGCGAAA 1203809  
 QY 227 AspThrValProCysTipSerPheGlyLeuAsnLysProAspPheLysAlaPheGlyLeu 246

DB 1203810 CATACCGTT-----TCTTTTGGGAAATAGTCGGAT-----TATTGGCTA 1203851  
 QY 247 IleGluGluAspGlyGlnLysTrpLeuAlaPheGlnPheAspLysLeuLeuProValGly 266  
 DB 1203852 AAACTGAAATGGCAAGCAATATTATGTTAAAGATGAAGTATTACCTTCTGAA 1203911  
 QY 267 GluLeuLysIleArgGlyAlaHisAsnTyrSerAsnAlaLeuAlaLeuAlaLeuGly 286  
 DB 1203912 GAAGCTACATTGTTGGTCCCATATATATGAACATTTTGGCAGCACACGATTTGGCA 1203971  
 QY 287 HisAlaValGlyLeuProPheAspAlaMetLeuGlyAlaLeuLysAlaPheSerGlyLeu 306  
 DB 1203972 CAAGCTATAGTATTAAATTTAGATTCAATTCGACGCACTTCGTCATTCAAAGGTTA 1204031  
 QY 307 AlaHisArgCysGlnTrpValArgGluArgGlnGlyValSerTyrTyrAspAspSerLys 326  
 DB 1204032 GATCATCGTTTCAATTTAGTCATCAAGCTAATGCGATTCGTTGGATTATGACTTAA 1204091  
 QY 327 AlaThrAsnValGlyAlaLeuAlaLeuAlaIleGluGlyLeuGlyAlaAspIleAspGly 346  
 DB 1204092 GCAACAAATGTGGGAGTACAGTTCGTCATTGGCTGGGCTT-----TATATTGAGGT 1204145  
 QY 347 LysLeuValLeuLeuAlaGlyGlyAspGlyLysGlyAlaAspPheHisAspLeuArgGlu 366  
 DB 1204146 AAATTGCATTTGTTGCTAGCGGAGACGGAAGGGCTGATTTTCAGAAATTAGCTGAA 1204205  
 QY 367 ProValAlaArg-----PheCysArgAlaValValLeuLeuGlyArgAsp 381  
 DB 1204206 TTAATTAATCAACCACACATTTATTGTTATGTT-----TTTGGTCGAGAT 1204250  
 QY 382 AlaGlyLeuIleAlaGlnAlaLeuGlyAsnAlaValProLeuValArgValAlaThrLeu 401  
 DB 1204251 GGTGCGCTGCTGCAAAATTTTCATCGCAAGCTTAT-----TTG 1204289  
 QY 402 AspGluAlaValArgGlnAlaAlaGluLeuAlaArg-----GluGlyAspAla 417  
 DB 1204290 TTCGATACANTGAACACGATAGTAATTTTACGCCCAACATTCGAAAGCGGAGATAG 1204349  
 QY 418 ValLeuLeuSerProAlaCysAlaSerLeuAspMetPheLysAsnPheGluArgGly 437  
 DB 1204350 GTATTATTGCGCTGCTGTCAGATCTGCAAGTCTCATGATTTGCTCTTTTCAAAAGCGCGC 1204409  
 QY 438 ArgLeuPhe 440  
 DB 1204410 GAAGAATTT 1204418  
 RESULT 7  
 US-09-252-991A-7928/G  
 ; Sequence 7928, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: 107196.136  
 ; CURRENT APPLICATION NUMBER: US/09/252,991A  
 ; CURRENT FILING DATE: 1999-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074,788  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/094,190  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 7928  
 ; LENGTH: 564  
 ; TYPE: DNA  
 ; ORGANISM: Pseudomonas aeruginosa  
 US-09-252-991A-7928  
 Alignment Scores:  
 Pred. No.: 2,01e-60 Length: 564  
 Score: 626.00 Matches: 127  
 Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 27.70% Indels: 0  
DB: 4 Gaps: 0

US-09-701-229-2 (1-448) x US-09-252-991A-7928 (1-564)

```
QY 322 TyrAspSerLysAlaThrAsnValGlyAlaAlaLeuAlaAlaLeuGluGlyLeuGly 341
D 563 TACGACGATTCACAGGCCACCAAGTCGCGCCCTTGGCGGCGATCGAGGGCTGGGT 504
QY 342 AlaAspLeuAspGlyLysLeuValLeuAlaGlyAlaGlyAspGlyLysGlyAlaAspPhe 361
D 503 GCCGACATCGACGCGAGCTGGTGTCTGCCCGCGGAGACGGCAAGGGCGCGATTC 444
QY 362 HisAspLeuArgGluProValAlaArgPheCysArgAlaValAlaValLeuGlyArgAsp 381
D 443 CATGACCTGCGGAGCGGTCGCGGCTTCTGCGGGGGTGGTACTGCTTGGCGGTCAC 384
QY 392 AlaGlyLeuLeuAlaGlnAlaLeuGlyAsnAlaValProLeuValArgValAlaThrLeu 401
D 383 GCCGGGCTGATTGCCAGGCACCTGGCAACGCGGTACGCTGGTGGCGCTGCCAACGCTG 324
QY 402 AspGluAlaValArgGlnAlaAlaGluLeuAlaArgGluGlyAspAlaValLeuLeuSer 421
D 323 GACGAAGCAGTCCGCGAGCGCGGAGCTGGCGCGGAGCGATGCGGCTGCTGTGTCG 264
QY 422 ProAlaCysAlaSerLeuAspMetPheLysAsnPheGluGluArgGlyArgLeuPheAla 441
D 263 CCGGCTCGCGGAGCGCTGGACATGTTCAAGAACTTCGAGAACCGGAGCGCTGTTCGCC 204
QY 442 LysAlaValGluGluLeuAla 448
D 203 AAAGCCGTAGAGGAGTAGCG 183
```

## RESULT 8

US-08-934-481-1

Sequence 1, Application US/08934481

Patent No. 5929045

GENERAL INFORMATION:

APPLICANT: Wallis, Nichola G.

APPLICANT: Fueyo, Joanna L.

APPLICANT: Lonetto, Michael A.

TITLE OF INVENTION: NOVEL MURD

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:

ADDRESSEE: Dechert Price & Rhoads

STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre

CITY: Philadelphia

STATE: PA

COUNTRY: US

ZIP: 19103

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/934.481

FILING DATE:

CLASSIFICATION: 514

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Dickinson, Todd O

REGISTRATION NUMBER: 28,354

REFERENCE/DOCKET NUMBER: GM10070A

TELEPHONE: 215-994-2252

TELEFAX: 215-994-2222

TELEX:

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 2193 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-934-481-1

## Alignment Scores:

Pred. No.: 5,08e-57 Length: 2193  
Score: 603.50 Matches: 164  
Percent Similarity: 53.16% Conservative: 80  
Best Local Similarity: 35.73% Mismatches: 178  
Query Match: 26.70% Indels: 37  
DB: 2 Gaps: 13

US-09-701-229-2 (1-448) x US-08-934-481-1 (1-2193)

```
QY 7 AspHisPheArg-----IleValValGlyLeuGlyLysSerGlyMetSerLeu 22
D 494 GATCAATTTAAATAATAAGAAAGTCTTGTGGTGGCCAGTCTGGTGAATCTGCA 553
QY 23 ValArgTyrLeuAlaAlaArgArgGlyLeuProPheAlaValAlaAspThrArg----- 39
D 554 GCTGCTTTGTTGGACAAGCTAGTGCCCATTTGTGACAGTAATATGATGGGAACCTTTCGAG 613
QY 40 GluAsnProProGluLeuAlaThrLeuArgAlaGlnTyrProGlnValGluValArgCys 59
D 614 GACAATCCAGCTGCCCAAAGTTTGTCTGAA-----GAAGGATCAAGTCAATTACA 664
QY 60 GlyGlu-----LeuAspAlaGluPheLeuCysSerAlaArgGluLeuTyr 74
D 665 GGTGCCCATCTCTTGGAACTCTTGGATGAAGAGTTTGTCTTATGGTGA----- 715
QY 75 ValSerProGlyLeuSerLeuArgThrProAlaLeuValGlnAlaAlaLysGlyVal 94
D 716 ---AATCCAGTATCCCTACAAATCCCATGATGAAAGGCTTTGCGCAAGGAAAT 772
QY 95 ArgIleSerGlyAspIleAspLeuPheAlaArgGluAlaLysAlaProIleValAlaLe 114
D 773 CCAGTCTTGACTGAGGTGGAATTTGCTTATTGATTTCAGAACGACCATGATTATG 832
QY 115 ThrGlySerAsnAlaLysSerThrValThrLeuValGlyGluMetAlaValAlaAla 134
D 833 ACAGATCGAAGCGTAAAGAACCAACCAACTATGATGGGGAAGTTTGTACTGCTGCT 892
QY 135 AspLysArgValAlaValGlyValAsnLeuGlyThrProAlaLeuAspLeu-----Leu 152
D 893 GGGCAACATGCTTTTATCAGGGAATATCGCTATCCAGTAGTCAGTAGTGGCCCAACT 952
QY 153 AlaAspAspIleGluLeuTyrValLeuGluLeuSerSerPheGlnLeuGluThrCysAsp 172
D 953 CGCTCGGACAAGGACACGCTTGTATGGAACCTTCTCTTCCAACTCATGCTGCTCAA 1012
QY 173 ArgLeuAsnAlaGluValAlaThrValLeuAsnValSerGluAspHisMetAspArgTyr 192
D 1013 GAATTCATCCAGAGATTGCGGTATTACCAACCTTATGCCAACTCATATGATGACAT 1072
QY 193 AspGlyMetAlaAspTyrHisLeuAlaLysHisArgIle-----PheArgGlyAla 209
D 1073 GGGTCATTTTCGGAATATGTAGCAGCCCAAGTGAATATCCAGAACAGATGACACAGCT 1132
QY 210 ArgGlnValValAlaAsnArgAlaAspAlaLeuThrArgProLeuIleAlaAspThrVal 229
D 1133 GATTTCTTGTCTTGAACCTTTAATCAAGACTTGTACTTCCAAAGACAGACCACTGTTGTA 1192
QY 230 ProCysTrpSerPheGlyLeuAsnLysProAspPheLysAlaPheGlyLeuIleGlu 249
D 1193 CCATTTTCAACA-----CTTGAAGAGTTTGTAT-----GGAGCTTATCTGAA 1234
QY 250 AspGlyGlnLysTrpLeuAlaPheGlnPheAspLysLeuLeuProValGlyGluLys 269
D 1235 GATGGTCAA-----CTCTACTTCCGTGTGAGTAGTATGTCGCGACGAATCAATCGT 1288
QY 270 IleArgGlyAlaHisAsnTyrSerAsnAlaLeuAlaLeuAlaLeuGlyHisAlaVal 289
```



```
Db 1289 GTTCCAGGTAGCCACATGTTGAAATGCCCTTGCAGACTATTCTGTAGCCAAAGCTTCGT 1348
QY 290 GlyLeuProPheAspAlaMetLeuGlyAlaLeuLysAlaPheSerGlyLeuAlaHisArg 309
Db 1349 GGTGTGGACATCAACCATCAAGGAACCTTTTCAGCCTTCGGTGTGCAACACCGCT 1408
QY 310 CysGlnTrpValArgGluArgGlnGlyValSerTyrTyrAspSerLysAlaThrAsn 329
Db 1409 CTCAGTGTGTGGATCAAGGCTGTTAAATCTATAACGACAGTAAATCAACTAAT 1468
QY 330 ValGlyAlaAlaLeuAlaLeuGlyAlaLeuGlyAlaAspGlyLysLeuVal 349
Db 1469 ATCTTGCTACTCAAAAGCCTTGTCCAGGATTGAC-----AACAGCAGGTGCTC 1519
QY 350 LeuLeuAlaGlyGlyAspGlyLysGlyAlaAspPheHisAspLeuArgGluProValAla 369
Db 1520 TTGATTGACAGTGTGTGGACCGTGCAGTGTGACGAATG---GTGCCAGACAT 1576
QY 370 ArgPheCysArgAlaValValLeuLeuGlyArgAspAlaGlyLeuLeuAlaGlnAlaLeu 389
Db 1577 ACTGGGACTCAAGAAGATGTCATCTGCTGCTCAACGCTGTCAGAACGTCGCAACGGCAGCA 1636
QY 390 GlyAsnAla---ValProLeuValArgValAlaThrLeuAspGluAlaValArgGlnAla 408
Db 1637 GACAAGCCTGGTGTCTGCTTATGTGGAGGACAGATATTGCAGATCCGACCCGCAAGGCC 1696
QY 409 AlaGluLeuAlaArgGluGlyAspAlaValLeuLeuSerProAlaCysAlaSerLeuAsp 428
Db 1697 TATGACCTGGCAGTCAAGGAGATGCTCTCTTCTAGTCTGCTGCTGCTGCTGCTGCTGCT 1756
QY 429 MetPheLysAsnPheGluGluArgGlyArgLeuPheAlaLysAlaValGluLeu 447
Db 1757 ATGTATGCTAACTTTGAAGTACGTGGCAGCTCTTTATCGACACAGTAGTGGGAGTTA 1813
```

## RESULT 9

```
US-09-290-602-1
; Sequence 1, Application US/09290602
; Patent No. 6350598
; GENERAL INFORMATION:
; APPLICANT: Wallis, Nicholas G.
; APPLICANT: Fuenyo, Joanna L.
; APPLICANT: Lonetto, Michael A.
; TITLE OF INVENTION: NOVEL MURD
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: US
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/290,602
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/934,481
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Dickinson, Todd O
; REGISTRATION NUMBER: 28,354
; REFERENCE/DOCKET NUMBER: GM10070A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-994-2252
; TELEFAX: 215-994-2222
; INFORMATION FOR SEQ ID NO: 1:
;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2193 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
;
; US-09-290-602-1
;
; Alignment Scores:
; Pred. No.: 5,08e-57 Length: 2193
; Score: 603.50 Matches: 164
; Percent Similarity: 53.16% Conservative: 80
; Best Local Similarity: 35.73% Mismatches: 178
; Query Match: 26.70% Indels: 37
; DB: 4 Gaps: 13
;
; US-09-701-229-2 (1-448) x US-09-290-602-1 (1-2193)
;
; QY 7 AspHisPheArg-----IleValValGlyLeuGlyLysSerGlyMetSerLeu 22
; Db 494 GATCAATTTAAATAAAGAAAGTCTTGTAGTTTGGCCAAAGTCTGTGAATCTGCA 553
; QY 23 ValArgTyrLeuAlaArgArgGlyLeuProPheAlaValValAspThrArg----- 39
; Db 554 GTCGTCTTGTGGACAGCTAGTGCCATGTGACAGTAATGATGGGAACCTTTCGAG 613
; QY 40 GluAsnProGluLeuAlaThrLeuArgAlaGlnTyrProGlnValGluValArgCys 59
; Db 614 GACAATCCAGCTGCCCAAGTTTGTCTGGAA-----GAAGGATCAAGGTCAATACA 664
; QY 60 GlyGlu-----LeuAspAlaGluPheLeuCysSerAlaArgGluLeuTyr 74
; Db 665 GTGGCCATCTTTGGAACTCTTGGATGAAGATTTGCTCTTATGTGTGAA----- 715
; QY 75 ValSerProGlyLeuSerLeuArgThrProAlaLeuValGlnAlaAlaLysGlyVal 94
; Db 716 ---ATCCAGGTATCCCTACAAATCCATGATTGAAGAGCTTTGCCAAGGAAT 772
; QY 95 ArgIleSerGlyAspIleAspLeuPheAlaArgGluAlaLysAlaProIleValAlaIle 114
; Db 773 CCAGTCTCTGCTGAGGTGGAATTTGCTTATTTGATTTCAGAACCCGATTTATGTATC 832
; QY 115 ThrGlySerAsnAlaLysSerThrValThrLeuValGlyGluMetAlaValAlaAla 134
; Db 833 ACAGGATCGAAGCGTAAGAACACCAACAGCTATGATTGGGGAAGTTTGGTCTGCT 892
; QY 135 AspLysArgValAlaValGlyGlyAsnLeuGlyThrProAlaLeuAspLeu----- 152
; Db 893 GGGCAACATGCTTTTATCAGGGAATATCGGCTATCCAGCTAGTCAAGTGGCCCAACT 952
; QY 153 AlaAspAspIleGluLeuTyrValLeuGluLeuSerSerPheGlnLeuGluThrCysAsp 172
; Db 953 GCGTCGGACAAAGGACACGCTTCTTATGGAACCTTCTCTTCCCACTCATGGTGTTCAA 1012
; QY 173 ArgLeuAsnAlaGluValAlaThrValLeuAsnValSerGluAspHisMetAspArgTyr 192
; Db 1013 GAATTTTCATCAGAGATTGCGGTTATTACCAACCTTATGCCAAGCTATATGACATACC 1072
; QY 193 AspGlyMetAlaAspTyrHisLeuAlaLysHisArgIle-----PheArgGlyAla 209
; Db 1073 GGGTCATTTTCGGAATATGTACGACCAAGTGGATATATCCAGAACAGATGACAGCAGCT 1132
; QY 210 ArgGlnValValAlaAsnArgAlaAspAlaLeuThrArgProLeuIleAlaAspThrVal 229
; Db 1133 GATTTTCCTTGTCTTGAACCTTTAATCAAGACTTTGACTTCCAGACAGAGCCACTGTGTA 1192
; QY 230 ProCysTrpSerPheGlyLeuAsnLysProAspPheLysAlaPheGlyLeuIleGluGlu 249
; Db 1193 CCATTTTCAACA-----CTTGAAGAGGTTGAT-----GGAGCTTATCTGGAA 1234
; QY 250 AspGlyGlnLysTrpLeuAlaPheGlnPheAspLysLeuLeuProValGlyGluLeuLys 269
; Db 1235 GATGTCAA-----CTCTACTTCTGGTGGTGAAGTAGTATGCGCAGCAAGTAATCGGT 1288
```





Db 2989 AATGAATCGGTGTTCCAGTAGCCACCAATGTGGAATATGCCCTTCGGACTATTGCTGTA 2930  
QY 286 GlyHisAlaValGlyLeuProPheAspAlaMetLeuGlyAlaLeuLysAlaPheSerGly 305  
Db 2929 GCCAAGCTTCGTGATGTGGACAATCAACCAATCAAGGAAATCTTTCAGCCCTTCGGTGT 2870  
QY 306 LeuAlaHisArgCysGlnTrpValArgGluArgGlnGlyValSerTyrTyrAspAspSer 325  
Db 2869 GTCAACACACCGCTCCAGTTGTGGATGACATCAAGGGTGTAAATCTTATACGACAGT 2810  
QY 326 LysAlaThrAsnValGlyAlaAlaLeuAlaAlaIleGluGlyLeuGlyAlaAspIleAsp 345  
Db 2809 AATCAACTATATCTTGGCTACTCAAAAGCCCTTCAGGATTTGAC-----AAC 2759  
QY 346 GlyLysLeuValLeuLeuAlaGlyAspGlyLysGlyAlaAspPheHisAspLeuArg 365  
Db 2758 AGCAAGTCGTCTGTGATGTGAGTGTGGACCGTGGCAATGAGTTTGACGAATG--- 2702  
QY 366 GluProValAlaArgPheCysArgAlaValValLeuLeuGlyArgAspAlaGlyLeuIle 385  
Db 2701 GTGCCAGACATTACTCGACTCAAGAAGATGGTCACTCGGTCAATCTGCAGAACGTGTC 2642  
QY 386 AlaGlnAlaLeuGlyAsnAla---ValProLeuValArgValAlaThrLeuAspGluAla 404  
Db 2641 AAACGGGACGACACAGGCTGTGCTGTATGTGAGCGGACACATATTCAGATGCG 2582  
QY 405 ValArgGlnAlaGluLeuAlaArgGluGlyAspAlaValLeuLeuSerProAlaCys 424  
Db 2581 ACCCGCAAGCCCTATCGACTTGGACTCAAGGAGATGTGTTCTTCTAGTCTGCCAAT 2522  
QY 425 AlaSerLeuAspMetPheLysAsnPheGluGluArgGlyArgLeuPheAlaValAla 444  
Db 2521 GCTAGCTGGGATATGATCTAACTTTGAAGTACGTGGGACCTCTTTATCGACACAGTA 2462  
QY 445 GluGluLeu 447  
Db 2461 GCGGAGTTA 2453

## RESULT 11

US-09-252-991A-7552  
; Sequence 7552, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 7552  
; LENGTH: 360  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-7552

Alignment Scores:  
Pred. No.: 7,73e-57 Length: 360  
Score: 591.00 Matches: 119  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 26.15% Indels: 0  
DB: 4 Gaps: 0

US-09-701-229-2 (1-448) x US-09-252-991A-7552 (1-360)

QY 5 AlaSerAspHisPheArgIleValValGlyLeuGlyLysSerGlyMetSerLeuValArg 24  
Db 3 GCCTCCGACCACTTCGGATCGTTGCGCCCTCGCAAGAGCGGCATGTCCTGCTGCGC 62

QY 25 TyrLeuAlaArgArgGlyLeuProPheAlaValAlaAspThrArgGluAsnProProGlu 44  
Db 63 TACCTGGCGCGCGCGCTTCCTTCGCGGTGTGATACCCGAGAGAACCCGCCGAG 122  
QY 45 LeuAlaThrLeuArgAlaGlnTyrProGlnValGluValArgCysGlyGluLeuAspAla 64  
Db 123 CTGGCCACCGCTGCGTCCAGTAGTATCCAGTGGAGTGGTTCGCGCGAATCGACGCC 182  
QY 65 GluPheLeuCysSerAlaArgGluLeuTyrValSerProGlyLeuSerLeuArgThrPro 84  
Db 183 GAGTTCTCTGCTCCCGCCGCAACTCTAGTCAAGCCCGCTGTGCTGCGCACCCCC 242  
QY 85 AlaLeuValGlnAlaAlaAlaLysGlyValArgIleSerGlyAspIleAspLeuPheAla 104  
Db 243 GCCTGTTACAGCGCGCGCGAAGGTGCGCATCTCCGTCGACATCGATCTTCGCC 302  
QY 105 ArgGluAlaLysAlaProIleValAlaIleThrGlySerAsnAlaLysSerThrVal 123  
Db 303 CGCGAGCGAAGCCCGCATGCTGCGCATCACCGGTTCCAAACGCGAAGAGCACCGTG 359

## RESULT 12

US-08-665-435A-1  
; Sequence 1, Application US/08665435A  
; Patent No. 5681694  
; GENERAL INFORMATION:  
; APPLICANT: Skatrud, Paul  
; APPLICANT: Peery, Robert  
; APPLICANT: Hoskins, JoAnn  
; APPLICANT: Wu, Chyun-yei Earnest  
; TITLE OF INVENTION: Biosynthetic Gene MurD of Streptococcus  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Eli Lilly and Company  
; STREET: Lilly Corporate Center  
; CITY: Indianapolis  
; STATE: Indiana  
; COUNTRY: US  
; ZIP: 46285  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/665,435A  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Webster, Thomas D  
; REGISTRATION NUMBER: 39,872  
; REFERENCE/DOCKET NUMBER: X-9900  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 317-276-3334  
; TELEFAX: 317-276-3861  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1350 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..1350  
US-08-665-435A-1

Alignment Scores:  
Pred. No.: 1.86e-55 Length: 1350  
Score: 596.50 Matches: 162

Percent Similarity: 52.92% Conservative: 83  
Best Local Similarity: 34.99% Mismatches: 177  
Query Match: 25.95% Indels: 41  
DB: 1 Gaps: 14

US-09-701-229-2 (1-448) x US-08-665-435A-1 (1-1350)

QY 7 AspHisPheArg-----ileValValGlyLeuGlySerGlyMetSerLeu 22  
DB 13 GATCAATTTAAATAAAGAGTTCTTTAGTTTGGCCCAAGCTCGTGAATCGCA 72  
QY 23 ValArgTyrLeuAlaArgGlyLeuProPheAlaValValAspThrArg----- 39  
DB 73 GCTCGTTTGGACAAGTAGTGCATTTGTGCACAGTAATGATGGAAACCTTTTCGAG 132  
QY 40 GluAsnProGluLeuAlaThrLeuArgAlaGlnTyrProGlnValGluValArgCys 59  
DB 133 GACATCCAGCTCCCAAGTTTGTGGAA-----GAGGGATCAAGGTCAATACA 183  
QY 60 GlyGlu-----LeuAspAlaGluPheLeuCysSerAlaArgGluLeuTyr 74  
DB 184 GGTGGCCATCTTTGGAACTCTTTGGATGAAGAGTTTGCCTTATGTGTA 234  
QY 75 ValSerProGlyLeuSerLeuArgThrProAlaLeuValGlnAlaAlaLysGlyVal 94  
DB 235 ---AATCCAGGTATCCCTCAACAATCCCATGATTGAAAGGCTTTGGCCAAAGAAAT 291  
QY 95 ArgIleSerGlyAspIleAspLeuPheAlaArgGluAlaLysAlaProIleValAlaIle 114  
DB 292 CCAGTCTGTAGTGGTGGTAAATTTGGCTTATTTGATTTGAGAACCCAGCATATTTGTATC 351  
QY 115 ThrGlySerAsnAlaLysSerThrValThrThrLeuValGlyGluMetAlaValAla 134  
DB 352 ACAGGATCGAAGCGTAAGCAACCAACCACTATGATGGGAAAGTTTGGACTGTGCT 411  
QY 135 AspIleArgValAlaValGlyGlyAsnLeuGlyThrProAlaLeuAspLeu-----Leu 152  
DB 412 GGCAACATGCTTTTATCAGGAATATCGGCTATCTCCAGTCCAGTGTGCTCAATA 471  
QY 153 AlaAspAspIleGluLeuTyrValLeuGluLeuSerPheGlnLeuGluThrCysAsp 172  
DB 472 GCATCAGATGAAGCACGCTGTGTATGGAATTTCTTTCCCACTCATGGTGTCAA 531  
QY 173 ArgLeuAsnAlaGluValAlaThrValLeuAsnValSerGluAspHisMetAspArgTyr 192  
DB 532 GAATTCATCCAGATGCGGTATATACCACTCATGCACTCATGCACTCATGCACTACCAT 591  
QY 193 AspGlyMetAlaAspTyrHisLeuAlaLysHisArgIle-----PheArgGlyAla 209  
DB 592 GGGTCAATTTCTGAATATGATGAGCCCAAGTGAATATCCAGAACAGATGACAGCAGCT 651  
QY 210 ArgGlnValValValAsnArgAlaAspAlaLeuThrArgProLeuIleAlaAspThr--- 228  
DB 652 GATTCCTGTCTGTGACTTTATCAAGACTTGGCAAAAGACTTGACTTCCACAGACAGAA 711  
QY 229 -----ValProCysTyrSerPheGlyLeuAsnLysProAspPheLysAlaPheGly 245  
DB 712 GCCACTGTTGTACCATTTTCAACA-----CTTGAAAGGTTGAT-----GGA 753  
QY 246 LeuIleGluAspGlyGlnLysTyrPheAlaPheGlnPheAspLysLeuLeuProVal 265  
DB 754 GCTTATCTGGAAGATGTCAC-----CTCTACTTCGTGTGTAAGTAGTCATGGCAGCG 807  
QY 266 GlyGluLeuLysIleArgGlyAlaHisAsnTyrSerAsnAlaLeuAlaLeuAlaLeu 285  
DB 808 AATGAATCGGTGTCCAGATGCCACAATGTGGAATGCTTGGCACTATTGCTGTA 867  
QY 286 GlyHisAlaValGlyLeuProPheAspAlaMetLeuGlyAlaLeuLysAlaPheSerGly 305  
DB 868 GCCAAGCTTGTGATGGACATCAACCATCAAGGAAACTCTTTGAGCTTCGGTGGT 927  
QY 306 LeuAlaHisArgCysGlnTyrValArgGluArgGlnGlyValSerTyrTyrAspSer 325  
DB 325

928 GTCAACACCCGCTCCAGTTTGTGGATGACATCAAGGGTGTAAATTCATATAACGACAGT 987  
QY 326 LysAlaThrAsnValGlyAlaAlaLeuAlaAlaIleGluGlyLeuGlyAlaAspIleAsp 345  
DB 988 AATCAACTATATCTTGGCTACTCAAAAGCCCTTATCAGGATTTGAC-----AAC 1038  
QY 346 GlyLysLeuValLeuLeuAlaGlyAspGlyLysGlyAlaAspPheHisAspLeuArg 365  
DB 1039 AGCAAGTCTGTCTGTATGTCAGGTGTTGGACCGTGGCAATGAGTTGACGAATTG--- 1095  
QY 366 GluProValAlaArgPheCysArgAlaValValLeuLeuGlyArgAspAlaGlyLeuIle 385  
DB 1096 GTGCCAGACATTAATGAGCTCAAGAGATGTCATCTGGGTCAATCTGCGAAGACGTGTC 1155  
QY 386 AlaGlnAlaLeuGlyAsnAla---ValProLeuValArgValAlaThrLeuAspGluAla 404  
DB 1156 AATCGGCGACGACACAAAGCTGTGTCTTATGTGGAGCGACAGATATTCAGATGCG 1215  
QY 405 ValArgGlnAlaAlaGluLeuAlaArgGluGlyAspAlaValLeuLeuSerProAlaCys 424  
DB 1216 ACCCGCAAGGCCCTATGAGCTTGGGACTCAAGGAGATGTGTTCTTCTTAGTCTGCAAT 1275  
QY 425 AlaSerLeuAspMetPheLysAsnPheGluGluArgGlyArgLeuPheAlaLysAlaVal 444  
DB 1276 GCCAGCTGGATATGATGCTAACTTTGAAAGTACGTGGCGACCTCTTTATCGACACAGTA 1335  
QY 445 GluGluLeu 447  
DB 1336 GCGGAGTTA 1344

RESULT 13  
US-08-665-435A-3  
Sequence 3, Application US/08665435A  
Patent No. 5681694  
GENERAL INFORMATION:  
APPLICANT: Skatrud, Paul  
APPLICANT: Peery, Robert  
APPLICANT: Hoskins, JoAnn  
APPLICANT: Wu, Chyun-Yeh Earnest  
TITLE OF INVENTION: Biosynthetic Gene MurD of Streptococcus  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Eli Lilly and Company  
STREET: Lilly Corporate Center  
CITY: Indianapolis  
STATE: Indiana  
COUNTRY: US  
ZIP: 46285  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/665,435A  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Webster, Thomas D  
REGISTRATION NUMBER: 39,872  
REFERENCE/DOCKET NUMBER: X-9900  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 317-276-3334  
TELEFAX: 317-276-3861  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1350 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: mRNA  
HYPOTHETICAL: NO

SEQUENCE :  
LENGTH :



```
QY 362 CCGCGAGCGGAGGCGCCCGATCGTGGCCATCACCGGTTCCAAACGCGAAGAGACACCGTGAC 421
Db 1027 CCGCGAGCGAAGGCGCCCGATCGTGGCCATCACCGGTTCCAAACGCGAAGAGACACCGTGAC 968
QY 422 CACCTGTGGCGGAAATGGCGGTGGCGCGGACAAAGCGTGTGCGCGTGGCGGGAACCT 481
Db 967 CACCTGTGGCGGAAATGGCGGTGGCGCGGACAAAGCGTGTGCGCGTGGCGGGAACCT 908
QY 482 CGGCAACCGCGGTGCGACCTGCTGGCGGAGACATCGAGCTGTACGTTGGAGCTGTC 541
Db 907 CGGCAACCGCGGTGCGACCTGCTGGCGGAGACATCGAGCTGTACGTTGGAGCTGTC 848
QY 542 GAGCTTCAGTGGAAACCTCGCATCGCTCAACCGCGAGGTGGCGACCTGCTGCAAGT 601
Db 847 GAGCTTCAGTGGAAACCTCGCATCGCTCAACCGCGAGGTGGCGACCTGCTGCAAGT 788
QY 602 CAGCGAAGACCATATGGATCGCTAGGAGCGCATGGCTGACTACCACTGGCGCAAGACCG 661
Db 787 CAGCGAAGACCATATGGATCGCTAGGAGCGCATGGCTGACTACCACTGGCGCAAGACCG 728
QY 662 GATCTTCGCGGTGCCCGCAGGTGCTGTAATCGCGCGCATGCCCTGACCCGACCGCT 721
Db 727 GATCTTCGCGGTGCCCGCAGGTGCTGTAATCGCGCGCATGCCCTGACCCGACCGCT 668
QY 722 GATCGCGATACCGTGGCGTCTGCTGCTGGCTGAAACGCGCTGAAAGCGGTTT 781
Db 667 GATCGCGATACCGTGGCGTCTGCTGCTGGCTGAAACGCGCTGAAAGCGGTTT 608
QY 782 CGGCTGATCGAGAGAGCGCCAGAGTGGCTGGCGGCAACATATTCCAACCGCTGCGCGCTGGC 841
Db 607 CGGCTGATCGAGAGAGCGCCAGAGTGGCTGGCGGCAACATATTCCAACCGCTGCGCGCTGGC 548
QY 842 GGTGGGCACTGAAGATCCGTGGCGGCGCAACATATTCCAACCGCTGCGCGCTGGC 901
Db 547 GGTGGGCACTGAAGATCCGTGGCGGCGCAACATATTCCAACCGCTGCGCGCTGGC 488
QY 902 GCTGGGCACTGGCTGGCGTGGCGTGGCGGCGATCGTGGCGGCTGAAGGCGTTTC 961
Db 487 GCTGGGCACTGGCTGGCGTGGCGTGGCGGCGATCGTGGCGGCTGAAGGCGTTTC 428
QY 962 CGGCTGGCTCATCGCTGCAAGTGGGTACGGAGCGCGGCGTGGCTGACTACGACGA 1021
Db 427 CGGCTGGCTCATCGCTGCAAGTGGGTACGGAGCGCGGCGTGGCTGACTACGACGA 368
QY 1022 TTCAAAGGCCAACACGTGGCGGCGCCCTGGCGGCGATCGAGGGGCTGGGTGCGGACAT 1081
Db 367 TTCAAAGGCCAACACGTGGCGGCGCCCTGGCGGCGATCGAGGGGCTGGGTGCGGACAT 308
QY 1082 CGAGCGCAAGCTGGTGTGCTGCGCGGCGGAGAGCGGCAAGGGCGCCGATTTCCATGACCT 1141
Db 307 CGAGCGCAAGCTGGTGTGCTGCGCGGCGGAGAGCGGCAAGGGCGCCGATTTCCATGACCT 248
QY 1142 CGCGAGCGGCTGCGCGCTTCTGCGCGGCGGTGGTACTGCTTGGCGGCTGACGCGCGGCT 1201
Db 247 CGCGAGCGGCTGCGCGCTTCTGCGCGGCGGTGGTACTGCTTGGCGGCTGACGCGCGGCT 188
QY 1202 GATTGCGGCACTGGGCAACCGGTTACCGCTGGTGGCGGCTGCGCAACGCTGGACGAAGC 1261
Db 187 GATTGCGGCACTGGGCAACCGGTTACCGCTGGTGGCGGCTGCGCAACGCTGGACGAAGC 128
QY 1262 AGTCGCGAGCGCGCGGCTGGCGCGGAGCGGATGCGGTGCTGCGCGGCGCTG 1321
Db 127 AGTCGCGAGCGCGCGGCTGGCGCGGAGCGGATGCGGTGCTGCGCGGCGCTG 68
QY 1322 CGCGAGCGCTGGACATGTTCAAGAACTTCGAAGAACGCGGAGCGGCTGTTCCGCCAAAGCGT 1381
Db 67 CGCGAGCGCTGGACATGTTCAAGAACTTCGAAGAACGCGGAGCGGCTGTTCCGCCAAAGCGT 8
QY 1382 AGAGGAG 1388
Db 7 AGAGGAG 1
```

```
RESULT 2
US-09-252-991A-7702
; Sequence 7702, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 7702
; LENGTH: 1371
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-7702

Query Match 94.18; Score 1364.6; DB 4; Length 1371;
Best Local Similarity 99.7%; Pred. No. 9.6e-249;
Matches 1367; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 27 AGCTGCGTTCGAGGACGAAGAGAGATGAGCTGATCGCTCCGACCACTTCCGCGATGTT 86
Db 1 AGCTGCGTTCGAGGACGAAGAGAGATGAGCTGATCGCTCCGACCACTTCCGCGATGTT 60
QY 87 GTCGGCTTCGCGAAGAGCGCATGTCCTGTGGCTTACCTGGCGCGCGCGGCTTCCCT 146
Db 61 GTCGGCTTCGCGAAGAGCGCATGTCCTGTGGCTTACCTGGCGCGCGCGGCTTCCCT 120
QY 147 TTGCGCGTTCGATACCCGAGAGAACCCCGCGGAGTGGCCACCTTGGCGGCGATAT 206
Db 121 TTGCGCGTTCGATACCCGAGAGAACCCCGCGGAGTGGCCACCTTGGCGGCGATAT 180
QY 207 CCGCAGGTGGAAGTGGCTTGGCGGAATCGAGCGGAGTTCCTCTCTCGCGCGGAA 266
Db 181 CCGCAGGTGGAAGTGGCTTGGCGGAATCGAGCGGAGTTCCTCTCTCGCGCGGAA 240
QY 267 CTCTATGTAGCCCCGGCTTGTGCTGCGCAACCCCTGCGTGTGTACAGGCGCGCGGAA 326
Db 241 CTCTATGTAGCCCCGGCTTGTGCTGCGCAACCCCTGCGTGTGTACAGGCGCGCGGAA 300
QY 327 GCGGTGCGCATCTCCGCTGACATCTCTTGGCGCGGAGGAGGCGCGGATGCTC 386
Db 301 GGTGTGCGCATCTCCGCTGACATCTCTTGGCGCGGAGGAGGCGCGGATGCTC 360
QY 387 GCCATCACCGGTTCCAAACGGAAGAGACACCGTGACCACTTGGTGGCGGAAATGGCGGTG 446
Db 361 GCCATCACCGGTTCCAAACGGAAGAGACACCGTGACCACTTGGTGGCGGAAATGGCGGTG 420
QY 447 GCGCGGACAAAGCGTGTGCGCGTGGCGGCAACCTTGGGACCCCGCGGCTGACCTGCTG 506
Db 421 GCGCGGACAAAGCGTGTGCGCGTGGCGGCAACCTTGGGACCCCGCGGCTGACCTGCTG 480
QY 507 GCGGACGACATCGAGCTGTACGTTGTGGAGCTGTCTGAGCTTCCAGCTTGGAAACCTTGGAT 566
Db 481 GCGGACGACATCGAGCTGTACGTTGTGGAGCTGTCTGAGCTTCCAGCTTGGAAACCTTGGAT 540
QY 567 CGCCTCAACCGCGGAGTGGCGACCGTGTCTGAAAGCTCAGCGAAGACCATATGATGATGCTAC 626
Db 541 CGCCTCAACCGCGGAGTGGCGACCGTGTCTGAAAGCTCAGCGAAGACCATATGATGATGCTAC 600
QY 627 GAGCGGATGGCTGACTACCACTGGCGAAGACCGGATCTTCCGCGGTGCCCGCAGGTC 686
Db 601 GAGCGGATGGCTGACTACCACTGGCGAAGACCGGATCTTCCGCGGTGCCCGCAGGTC 660
QY 687 GTGGTGAATCGCGCGGATGCGCTGACCGACCGCTGATCGCGGATACCGTGGCGGCTGCTGG 746
Db 661 GTGGTGAATCGCGCGGATGCGCTGACCGACCGCTGATCGCGGATACCGTGGCGGCTGCTGG 720
```

Query Match 39.1%; Score 567; DB 4: Length 567;  
Best Local Similarity 100.0%; Pred. No. 1.3e-98;  
Matches 567; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 772 TCAAGGCTTTTCGCCCTGATCGAGGAACACGGCCAGAGTGGCTGGCGTTCACAGTTTCGACA 831

## RESULT 4

	Query Match	30.2%; Score 438; DB 4; Length 564;
	Best Local Similarity	100.0%; Pred. No. 2.4e-74;
	Matches 438; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
Qy	1013 CTAGGAGATTCAAAGGCCACCAACGTCTGGCGCGCCCTTGGCGGCATCAGGGGCTGGG	1072
Dd	564 CTACGAGATTCAAAGGCCACCACAGTCGCGCGCCCTTGCGCGCATCAGGGGCTGGG	505
Qy	1073 TGCGGCATCGACGGCAAGCTGGTGTCTGCTCGCGGGGAGACGGCAAGGGCGCCGATTT	1132
Dd	504 TGCGGCATCGACGGCAAGCTGGTGTCTGCTCGCGGGGAGACGGCAAGGGCGCCGATTT	445



QY 1133 CCATGACCTGCGGAGCGCGTTCGCGCGCTTCTGCGGCGGTGCTACTGCTTGGCCGTGA 1192  
|||||  
Db 444 CCATGACCTGCGGAGCGGTGCGCGCTTCTGCGGCGGTGCTACTGCTTGGCCGTGA 385  
QY 1193 GCCGGGCTGATGTCGCCAGGCACTGGGCAACGCGGTACCGCTGTCGCGTGGCAACGCT 1252  
|||||  
Db 384 GCCGGGCTGATGTCGCCAGGCACTGGGCAACGCGGTACCGCTGTCGCGTGGCAACGCT 325  
QY 1253 GGACGAAGCACTCCGGCAGGCGCGGAGCTGGCGCGCGGCAAGCGATGCGGTGCTGTC 1312  
|||||  
Db 324 GGACGAAGCACTCCGGCAGGCGCGGAGCTGGCGCGCGGCAAGCGATGCGGTGCTGTC 265  
QY 1313 GCCGGCCTGCGGAGCGCTTGCATGTTTCAAGAACTTCGAAGAACCGGACCGCTTTCGC 1372  
|||||  
Db 264 GCCGGCCTGCGGAGCGCTTGCATGTTTCAAGAACTTCGAAGAACCGGACCGCTTTCGC 205  
QY 1373 CAAGCCGTAGAGGAGCTAGCGTATGCTGTCGCTGTTGCGCCCTTCCCTGCGCGCTG 1432  
|||||  
Db 204 CAAGCCGTAGAGGAGCTAGCGTATGCTGTCGCTGTTGCGCCCTTCCCTGCGCGCTG 145  
QY 1433 TTGAGCCGCGCACGCATC 1450  
|||||  
Db 144 TTGAGCCGCGCACGCATC 127

RESULT 5  
US-09-252-991A-7552  
; Sequence 7552, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252.991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 7552  
; LENGTH: 360  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-7552

Query Match 24.4%; Score 353.6; DB 4; Length 360;  
Best Local Similarity 98.9%; Pred. No. 1.8e-58;  
Matches 356; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 61 TCGCCTCCGACCACTTCGCGATCTGTCGGCCTCGGCAAGCGGATGTCCTGCTGTC 120  
|||||  
Db 1 TCGCCTCCGACCACTTCGCGATCTGTCGGCCTCGGCAAGCGGATGTCCTGCTGTC 60  
QY 121 GCTACTGCGCGCGCGGCTGCTTTCGCGCTGTCGATACCCGAGAACCCCGCG 180  
|||||  
Db 61 GCTACTGCGCGCGCGGCTGCTTTCGCGCTGTCGATACCCGAGAACCCCGCG 120  
QY 181 AGCTGGCCACCTCCGTCGCCAGTATCCGAGGTGGAAGTCCGTCGTCGCGACG 240  
|||||  
Db 121 AGCTGGCCACCTCCGTCGCCAGTATCCGAGGTGGAAGTCCGTCGTCGCGACG 180  
QY 241 CCGAGTTCCTCTGTCGCCCGCGGAACTATGTCAGCCCGGCTTTCGTCGCGACCC 300  
|||||  
Db 181 CCGAGTTCCTCTGTCGCCCGCGGAACTATGTCAGCCCGGCTTTCGTCGCGACCC 240  
QY 301 CTCGCTGGTACAGCGCGCGGAAAGCGTGGCGCATCTCCGCTGACATCATCTCTTCG 360  
|||||  
Db 241 CCGCGCTGGTACAGCGCGCGGAAAGGTGTGCGCATCTCCGCTGACATCATCTCTTCG 300  
QY 361 CCGCGGAGGCGAAGGCCCGCATGTCGCCATCACCCTTCACAGCGGCAAGCACCGTGA 420

|||||  
Db 301 CCGCGGAGGCGAAGGCCCGCATGTCGCCATCACCCTTCACAGCGGCAAGCACCGTGA 360  
RESULT 6  
US-09-252-991A-7624  
; Sequence 7624, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252.991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 7624  
; LENGTH: 1359  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-7624

Query Match 17.8%; Score 258; DB 4; Length 1359;  
Best Local Similarity 100.0%; Pred. No. 1.9e-40;  
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1193 CGCGGGCTGATTCGCCAGGCACTGGGCAACGCGGTACCGCTGTCGCGTGGCAACGCT 1252  
|||||  
Db 1 CGCGGGCTGATTCGCCAGGCACTGGGCAACGCGGTACCGCTGTCGCGTGGCAACGCT 60  
QY 1253 GGACGAAGCACTCCGGCAGGCGCGGAGCTGGCGCGGCAAGCGATGCGGTGCTGTC 1312  
|||||  
Db 61 GGACGAAGCACTCCGGCAGGCGCGGAGCTGGCGCGGCAAGCGATGCGGTGCTGTC 120  
QY 1313 CGCGGCTGCGGAGCGCTTGCATGTTTCAAGAACTTCGAAGAACCGGACCGCTTTCGC 1372  
|||||  
Db 121 CGCGGCTGCGGAGCGCTTGCATGTTTCAAGAACTTCGAAGAACCGGACCGCTTTCGC 180  
QY 1373 CAAGCCGTAGAGGAGCTAGCGTATGCTGTCGCTGTTGCGCCCTTCCCTGCGCGCTG 1432  
|||||  
Db 181 CAAGCCGTAGAGGAGCTAGCGTATGCTGTCGCTGTTGCGCCCTTCCCTGCGCGCTG 240  
QY 1433 TTGAGCCGCGCACGCATC 1450  
|||||  
Db 241 TTGAGCCGCGCACGCATC 258

RESULT 7  
US-09-328-352-1282  
; Sequence 1282, Application US/09328352  
; Patent No. 6562958  
; GENERAL INFORMATION:  
; APPLICANT: Gary L. Breton et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
; FILE REFERENCE: GTC99-03PA  
; CURRENT APPLICATION NUMBER: US/09/328.352  
; CURRENT FILING DATE: 1999-06-04  
; NUMBER OF SEQ ID NOS: 8252  
; SEQ ID NO 1282  
; LENGTH: 1416  
; TYPE: DNA  
; ORGANISM: Acinetobacter baumannii  
US-09-328-352-1282

Query Match 16.5%; Score 239; DB 4; Length 1416;  
Best Local Similarity 51.6%; Pred. No. 7.3e-37;  
Matches 604; Conservative 0; Mismatches 555; Indels 12; Gaps 2;



Db 1300 GAAACAAAGCTGAAGATGGTATGCTCATCACCAGCATGTCAAGTTTGTGATGTTT 1359  
 QY 1341 AAGAACTTCGAAGAAGCGGACGCGCTCTTCG 1371  
 Db 1360 AAAAGTTATATGACCGTGGTCAGCAGTTTG 1390

RESULT 8  
 US-09-103-840A-1/c  
 ; Sequence 1, Application US/09103840A  
 ; Patent No. 6294328  
 ; GENERAL INFORMATION:  
 ; APPLICANT: FLEISCHMAN, Robert D.  
 ; APPLICANT: WHITE, Owen R.  
 ; APPLICANT: FRASER, Claire M.  
 ; APPLICANT: VENTER, John C.  
 ; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
 ; TITLE OF INVENTION: TUBERCULOSIS  
 ; FILE REFERENCE: 24366-20007-00  
 ; CURRENT APPLICATION NUMBER: US/09/103,840A  
 ; CURRENT FILING DATE: 1998-06-24  
 ; NUMBER OF SEQ ID NOS: 2  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 1  
 ; LENGTH: 4411529  
 ; TYPE: DNA  
 ; ORGANISM: Mycobacterium tuberculosis  
 ; OTHER INFORMATION: H37Rv  
 US-09-103-840A-1

Query Match 12.0%; Score 173.8; DB 3; Length 4411529;  
 Best Local Similarity 51.7%; Pred. No. 2.3e-24;  
 Matches 499; Conservative 0; Mismatches 427; Indels 39; Gaps 3;

QY 273 GTCAGCCCCGGCTTGCTGCTGCGCACCCCTGCGCTGGTACAGGCCCGCCCGAAGCGGTG 332  
 Db 2416179 GCCAGTCCCGGCTTCGCGCCGCAAGCCCGCTACTGCGCGCGCGCCGCGCGGGGTG 241616  
 QY 333 CGCATCTCCGCTGCATCATGATCTCTTCGCCG-----CGAGGCG 371  
 Db 2416119 CCAGTCTGGGGTACGTGGAGTTAGGCTTGGCGGCTAGACGACGCGGCTGCTACGGACCG 2416160  
 QY 372 AAGCGCCCGATCTGCCCATCACCGGTTCACAGCGGAAGACACCGTGTACCCACCTCGGTG 431  
 Db 2416059 CCGCGAGCTGGCTGGTGTGACCGCACCAAGCGCAAGACACACGACGCTCGATGCTG 2416058  
 QY 432 GCGGAATGCGGTGGCGGGGACAGAGTGTGCGCGTGGCGGGCAACCTCGGCACCCCG 491  
 Db 2415999 CAGCGCATGTGATCGCGGTGGCGCGCGCGCTGCTGTGGCGCAATATCGSCAGTGCG 2415998  
 QY 492 GCGCTGCACCTGTGCGCGAGGACATCGAGCTGTACGTGTGGAGCTGTGAGCTTCCAG 551  
 Db 2415939 GTGCTGGATGTGCTGACGAGCGCGCGAGCTCTGCGCGTGGAGTGTCCAGTTCCAG 2415938  
 QY 552 CTGGAACCTTGCATCGCCTCAACGCGGAGGTGGCGACCGTGTGTAACCTGACCGAAGAC 611  
 Db 2415879 CTGCACTGGGCGCGCTGCTGCGGCCCGAGCGCGCGTGTCTCAACATTCGCGAAGAC 2415878  
 QY 612 CATATGATGCTGACGAGCGGATGGCTGACTACCACTGGCGAAGCACCGGATCTTCGCG 671  
 Db 2415819 CACCTGGACTGGCATGCCACGATGGCGGAATACACGCGGCGCAAGCCCGGTGTGACC 2415818  
 QY 672 GGTGCCCGCAGTCTGTGTAATCGCGCGATGCCCTGACCCGACCGTGTATCGCGAT 731  
 Db 2415759 GCGGGGTGACGTGTGCGCGGTGGATGACGCGAGCGCGCGCATGCTGTGACGCTCA 2415758  
 QY 732 ACCGTCCGTGTGTTGTCG---TTCGGCTTGAACAGCCCGGACTTCAAGCTTTCGCGCTG 788  
 Db 2415699 CCGGCGCAGGTGCGGGTTCGGCTTCGGGCTCGGGAGCGCGCGCGGGAATCGGCGTG 2415698  
 QY 789 ATCGAGGAAGCGGCGAGAGTGGTGGCGTTCCAGTTTCGAAGCTGTCTCGCGTTCG 848

Db	2415639	CGGACGCCCACTGGTGCATCGCGGCTTCTCCGACGACTTGACGTGTGCGGTCGCG	2415580
QY	849	GAACTGAAGATCCGTGGCGCCACAACTATTTCAACGCGCTCGCGCGCTGCGCTGCGGC	908
Db	2415579	TCGATACCGGTGCCAGTCCGGTCGCGTGCATTGACGCCCTGGCGCGCGCGCTGGCC	2415520
QY	909	CATGCGGTGGCTCCGCTTTCGACGCCATGCTGGCGCGCTGAAGCGGTTCCTCGGCGCTG	968
Db	2415519	CGCTCGGTGGGTGCCCGCGTGCATCGCGACGCGCTCACGCTCTTCAGTGGGC	2415460
QY	969	GCTCATCTGCTGCCAGTGGGTACGCGAGCGGCAGGGCGTGAAGTACTACGACGATCCAA	1028
Db	2415459	CGACACCGCGCGAGTGTGTGCCGTTCGCGAGCGCATCACCTACGTGACGACTCCA	2415400
QY	1029	GCCACCAAGTCCGGCGCCGCCCTTGGCGGGGATCGAGGGGTGGGTGCCACATCGACGC	1088
Db	2415399	GCCACCAACCGCACGCCGCGGGGCTTCGGTGCTTG-----CATACCG	2415355
QY	1089	AAGCTGGTGTGCTCCGCGGGGAGACGGCAGGGCGCGGATTTCCATGACCTGCCGAG	1148
Db	2415354	AGGGTGGTATGGATCCCGGTGGCTTGTCTCAAGGGCGCGTCGCTTCAACCGAGGTGGC	2415295
QY	1149	CCGGTCCGCGCTTCTGCCCGGGGGTGTACTGCTTGGCCGTGACGCCCGGCTGATTGCC	1208
Db	2415294	GCGATGGCTCGCGGCTGTGCTGGCGGTGTGATCGGCCGGATCCGACGGTTC	2415235
QY	1209	CAGGC 1213	
Db	2415234	GAGGC 2415230	

## RESULT 9

```

US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6394328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

```

	Query Match	11.9%	Score 172.2;	DB 3;	Length 4403765;	
	Best Local Similarity	51.6%;	Pred. No. 4..6e-24;			
	Matches 498;	Conservative 0;	Mismatches 428;	Indels 39;	Gaps 3;	
QY	273	GTACGCCCGGGTTGCTGC	GTGCACCCCTGGCGTG	TACAGCGCGCCGGAAGCGTGG	332	
Db	2414838	GGCAGTCGCGGGTTCTCG	CCCCCAATCCCGCTACT	GGCGCGCGCGCGGGGTG	2414779	
QY	333	CGCATCTCCGGTGAACTC	GATCTCTTCGCCCG-	-----CGAGCGG	371	
Db	2414778	CCGATCTGGGGTGACGTG	GAGTTAGCTTGGCGGT	TAGACGCAGCGGGTGCT	ACGACCG	2414719
QY	372	ANGGCCCGCATCGTGC	CCATCACCGGTTC	CAACGGGAAGACACCGT	TGACCACCGTGGTG	431
Db	2414718	CGCGCGAGCTGGCTGGT	GTGTACCGGCACCAAC	CGGCAACGACACCAAC	CACGACGTCGATGCTG	2414659

```

RESULT 10
US-09-557-884-1
; Sequence 1, Application US/09557884
; Patent No. 6506581
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: The Nucleotide sequence of
; the Haemophilus influenzae Rd Genome, Fragments
; Thereof, and Uses Thereof
;
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: MD
; COUNTRY: USA

```

APPLICANT: Fleischmann et al.  
TITLE OF INVENTION: The Nucleotide sequence of the Haemophilus influenzae Rd Genome, Fragments Thereof, and Uses Thereof

ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3 1/2 inch diskette  
COMPUTER: Dell Pentium  
OPERATING SYSTEM: MS DOS v6.22  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/557,884  
FILING DATE: 25-Apr-2000  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/476,102  
FILING DATE: JUN-5-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Michelle S. Marks  
REGISTRATION NUMBER: 41,971  
REFERENCE/DOCKET NUMBER: PB186P3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 301-309-8504  
TELEFAX: 301-309-8439  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1830121 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
US-09-557-884-1  
Query Match 10.3%; Score 149.8; DB 4; Length 1830121;  
Best Local Similarity 51.0%; Pred. No. 7.2e-20;  
Matches 440; Conservative 0; Mismatches 407; Indels 16; Gaps 3;  
QY 272 TGTACGCCCGGCTGTGCTGCGCACCCCTCGCTGGTACAGGCGCGCGAAAGGCT 331  
DB 1203341 TATTAGCCAGGGCTTGGGTAACCAACCAAAATTCACACCGCACITAAAGCGGACT 1203400  
QY 332 GCGCATCTCGGTGACATGATCTCTTCGCCGCGAGGAGGCGCGCATCGTCCGCAT 391  
DB 1203401 GGAAGTAATCGCGCATATTGAATATTCTGCGCGCAGGACAAAGCCAAATTTGTGGGAT 1203460  
QY 392 CACCGGCTTCAACGCGAAGAGCAGCGTGACACACCTGTTGGCGGAAATGGCGGTGCGCGC 451  
DB 1203461 TACAGGTTCAATGTAAAGTACCGTACTACTTGTAGTTATGAAATGGCGAAAGCTGC 1203520  
QY 452 GGACAAGCGTGTCCGCTGCGCGCAACCTCGSCACCCGCGCTCGACCTGCTGCGCGCA 511  
DB 1203521 TGGTGTGAAAGTTGTTATGGCGGAAATATTGGGATTCGCCGCTTGTCTATTCTTGAATGA 1203580  
QY 512 CGACATCGAGCTGTAGCTGTGGAGCTGTGAGCTTCAGCTGGAACCTGCGATCGCGCT 571  
DB 1203581 AGATTGTGAACCTTTATGACTAGAGCTTTCTAGTTTTCAGCTTGAGACAACCTTATAGCTT 1203640  
QY 572 CAACGCGAGGTGGCGCGCTGTAACCTCAGCGAAGACCATATGATCGTACGACG 631  
DB 1203641 AAAAGTGGCGGAGCGACTGTCTTGAACCTGACTGAAGATCATATGATGATCGTATATGA 1203700  
QY 632 CATGGCTGACTACCACTGGCCCAAGCACCGGATCTTCGCGGCTGCGCGCGGCGGCTGCTGT 691  
DB 1203701 TTTAGAAGATTATCCCAAGCAAAATTAGCATTTATCATATGCTAAAGTAGTGTGTT 1203760  
QY 692 GAATCGCGCGATGCGCTGACCCGACCGCTGATTCGCGGAT--ACCGTGGCGGTGCTGGTGC 749  
DB 1203761 GAACATGAAGATAGCTGACTTTTGGGGAACCAAAATCAAGCGAAACATACCGTTTC 1203820  
QY 750 TTCGCGCTGAACAACCGGACTTCAAGGCTTTCGCGCTCATCGAGGAACGCGCGCAGAG 809  
DB 1203821 TTTTCGGAATAATAGTCGGGATTTATGGCTAAAACT-----GAAATGGCAAGCAA 1203872  
QY 810 TGGCTGGCGTTCCAGTTCACAAAGCTGTGCGGTTGGCGAACTGAAGATCGTGGCGCGC 869  
DB 1203873 TATTAAATGGTAAGATGAAGTGAATTTTACCTTGTGAGAGCTACATGTTGGTGGTCC 1203932

QY 870 CACAACATATCCAAACGCGCTCGCGGCTGGCGCTGGCGCATCGGCTCGGCTCGCGCTTC 929  
DB 1203933 CATAATTATATGAACATTTTGGCAGCAACAGCATTTGGCACAAGCTATAGTATTATTTA 1203992  
QY 930 GAGCCATCTCGCGCGCTGAGGCGTTTTCGGGCTCGGCTATCGCTGCCAGTGGTA 989  
DB 1203993 GATTCAATTCGTACCGCACTTCGTCAATTTCAAAGGGTTAGATCATCGTTTCAATTAGTG 1204052  
QY 990 CCGAGCGCGCAGGCGTGAGCTACTACGACGATTTCGAAGGCCACCAACGTCGCGCGCGCC 1049  
DB 1204053 CATCAAGCTAATGGCATTCGTTGGATTAACTCTAAAGCAACAATGTGGGAGTACA 1204112  
QY 1050 CTGGCGGCGATCGAGGGGCTGGGTGCCGACATCGACGAGCAAGCTGTGTCTCGCGCGC 1109  
DB 1204113 GTTGTGCTATGCTGGGCT-----TTATATTAGGGTAAATTCATTTGTTGCTAGGC 1204166  
QY 1110 GGAGCGGCAAGGGCGCGGATTT 1132  
DB 1204167 GGAGACGGAAAGGGGCTGATTT 1204189  
RESULT 11  
US-09-643-990A-1  
; Sequence 1, Application US/09643990A  
; Patent No. 6528289  
; GENERAL INFORMATION:  
; APPLICANT: Robert D. Fleischmann  
; Mark D. Adams  
; Owen White  
; Hamilton O. Smith  
; J. Craig Venter  
; TITLE OF INVENTION: The Nucleotide sequence of  
; the Haemophilus Influenzae Rd Genome, Fragments  
; Thereof, and Uses Thereof  
; NUMBER OF SEQUENCES: 1  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville,  
; STATE: MD  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3 1/2 inch diskette  
; COMPUTER: Dell Pentium  
; OPERATING SYSTEM: MS DOS v6.22  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/643,990A  
; FILING DATE: 23-Aug-2000  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/487,429  
; FILING DATE: 1995-06-07  
; APPLICATION NUMBER: 08/426,787  
; FILING DATE: 1995-04-21  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Kenley K. Hoover  
; REGISTRATION NUMBER: 40,302  
; REFERENCE/DOCKET NUMBER: PB186PIC1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 301-610-5790  
; TELEFAX: 310-309-8439  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1830121 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
US-09-643-990A-1

Query Match

10.3%; Score 149.8; DB 4; Length 1830121;

```

Best Local Similarity 51.0%; Pred. No. 7.2e-20;
Matches 440; Conservative 0; Mismatches 407; Indels 16; Gaps 3;

QY 272 TGTACGCGCGGCTGTGCTGCGCACCCCTGCGCTGTGACAGCCGCGCGAAAGGGCT 331
DB 1203341 TATTAGCCAGGGCTTGGGTAAACACACAGAAATTAACACCGACTTAAGCGGGAGT 1203400
QY 332 CGGCATCTCCGGTGACATCGATCTCTTCGCGCGAGCGAGGCCCGCATCGTCGCCAT 391
DB 1203401 GGAAGTAATCGCGATATTGAATTATCTGCGCGCAGCAGCAAGCAATTTGGGGAT 1203460
QY 392 CACCGGTTCCAAACGGAAGACACCGGTGACACCCCTGTGTGGCGGAAATGGCGGTGGCGC 451
DB 1203461 TACAGGTTCAATGGTAAAGTACCGTAACCTTATAGTTTATGAAATGGCGAAAGCTGC 1203520
QY 452 GGACAAGGCTGTGCGCGTGGCGGCGCAACCTCGGACCCCGCGCTGCAGCTGCTGGCGGA 511
DB 1203521 TGGTGTGAAAGTTGGTATGGCGGGAATATTGGGATTCGCCCTTTGTCATTTGAATGA 1203580
QY 512 CGACATCGAGCTGTACGTTGGAGCTGTGAGCTTCCAGCTGGAACCTCGCATCGCCT 571
DB 1203581 AGATTGTGAACCTTTATGTACTAGAGCTTTCTAGTTTTCAGCTTGAGACAACTTATAGCTT 1203640
QY 572 CAACGCGGAGTGGCGGCGGCTGTGAGCTGAGCGAAGACCATATGATGATCGCTACGACGG 631
DB 1203641 AAAAGCTGCGCGCAGCGACTGTCTTGAAGCTGACTGAAGATCATATGGATCGCTATATGA 1203700
QY 632 CATGGCTGACTPACCACTGGCGCAAGCAGCGGATCTTCGCGGTGCGCGCCAGGTCGTGGT 691
DB 1203701 TTTAGAGATTTATGCCAAGCAAAATTTACGATTTTATCATATGATGATGATGTTGTT 1203760
QY 692 GAATCGGCGGATCCCTGTGACCGCGGCTGTGATCGCGAT--ACCGTGGCTGTGCTGTCG 749
DB 1203761 GAACAATGAAGATAGGCTGACTTTTGGGAAACGAAATCAACGAAACATACCGTTTC 1203820
QY 750 TTCGCGCTGACACGCGGACTTCAAGCTTTCGCGCTGTGATCGAGGAGACGCGCGCAGAAG 809
DB 1203821 TTTTGGGAAATAGTCGGGATTTGCGCTAAAAACT-----GAAATGGCAAGCAA 1203872
QY 810 TGGCTGGCGTTCACAGTTCGACAACTGCTGCGGTTGGCGAATGGAAGTCCGTCGCGCC 869
DB 1203873 TATTATGTAAGATGAGTATGATTTACCTTGTGAAGAGTACATGTTGGTGGTCG 1203932
QY 870 CACAATATTCCAAAGCGCTGCGCGCTGCGCGCTGGCGCATGCGGCTGCGCTCCCGTTC 929
DB 1203933 CATAATTATATGACATTTTGGCAGCAACAGCATTTGCACAAGCTATAGTATTAATTA 1203992
QY 930 GACGCCATGCTCGCGCGCTGAAGCGCTTTTCGCGCTGCTCATCCTGCCAGTGGTA 989
DB 1203993 GATTCATTCGTACCGCACCTGTCATTTCAAAGGGTTAGATCATCTTTTCAATTAGT 1204052
QY 990 CGCGAGCGGCGGCGGTGAGCTACTAGCAGGATTCACAGGCGCAACGTCGCGCGCGCC 1049
DB 1204053 CATCAAGCTAATGCACTTCGTTGATTAATGACTTAAGCAACAATTTGGGAGTACA 1204112
QY 1050 CTGCGCGGATGAGGGCGTGGTGGCGGATGCGGATGCGGAGTGGTGTGCTGCGCGCGC 1109
DB 1204113 GTTCTGCTATTTGCTGGGCT-----TTATATTGAGGTAATTTGCTATTGCTAGGC 1204166
QY 1110 GGAGACGGAAGGCGCGGATTT 1132
DB 1204167 GGAGACGGAAGGCGCTGATTT 1204189

```

```

RESULT 12
US-09-252-991A-7929/c
; Sequence 7929, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136

```

```

; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; LENGTH: 1170
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-7929

Query Match      8.3%; Score 120; DB 4; Length 1170;
Best Local Similarity 100.0%; Pred. No. 1.8e-14;
Matches 120; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GTGCTGATCGCGCTCGCCACCTTGAAGCTGCGTTGAGGACGAAGAGCATGAGCCTGAT 61
DB 120 GTGCTGATCGCGCTCGCCACCTTGAAGCTGCGTTGAGGACGAAGAGCATGAGCCTGAT 61
QY 62 CGCCTCCGACCATCTCCCATCGTTGCGGCTCGGCAAGAGCGGCATGTCCTGGTGGC 121
DB 60 CGCCTCCGACCATCTCCCATCGTTGCGGCTCGGCAAGAGCGGCATGTCCTGGTGGC 1

RESULT 13
US-08-920-812-13
; Sequence 13, Application US/08920812
; Patent No. 5763188
; GENERAL INFORMATION:
; APPLICANT: Ohno, Tsuneoya
; APPLICANT: Matsuhisa, Akio
; APPLICANT: Uehara, Hirotsugu
; APPLICANT: Eda, Soji
; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/920,812
; FILING DATE: 29-AUG-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/362,577
; FILING DATE: 27-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Rin-Laures, Li-Hsien
; REGISTRATION NUMBER: 33,547
; REFERENCE/DOCKET NUMBER: 19036/32420
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9515 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Pseudomonas aeruginosa

```

[illegible]





**THIS PAGE BLANK (USPTO)**





; NAME/KEY: CDS									
; LOCATION: (1)...(1347)									
US-09-815-242-7912									
Query Match 92.9%; Score 1347; DB 9; Length 1347;									
Best Local Similarity 100.0%; Pred. No. 0;									
Matches 1347; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
QY	51	ATGAGCCTGATCGCCTCCGACCACTTCGCGATCGTTGTGCGCCCTCGGCAAGAGCGCATG	110						
DB	1	ATGAGCCTGATCGCCTCCGACCACTTCGCGATCGTTGTGCGCCCTCGGCAAGAGCGCATG	60						
QY	111	TCCTGTGGTGCTACCTGGCGGCCCGCGGCTTGCCCTTCGCGGTGTCGATACCCGAGAG	170						
DB	61	TCCTGTGGTGCTACCTGGCGGCCCGCGGCTTGCCCTTCGCGGTGTCGATACCCGAGAG	120						
QY	171	ACCCTCCGAGCTGGCCACCTTCGCTGCGCCGCGGAACTCTATGTCAAGCCCGGCTTGCGG	230						
DB	121	ACCCTCCGAGCTGGCCACCTTCGCTGCGCCGCGGAACTCTATGTCAAGCCCGGCTTGCGG	180						
QY	231	GAATCGAGCGCGAGTTCTCTGCTCGCGCGGAACTCTATGTCAAGCCCGGCTTGCGG	290						
DB	181	GAATCGAGCGCGAGTTCTCTGCTCGCGCGGAACTCTATGTCAAGCCCGGCTTGCGG	240						
QY	291	CTGCGCACCCCTGCGCTGTGTACAGCGCCCGCGGAAAGCGTGGCATCTCCGCTGACATC	350						
DB	241	CTGCGCACCCCTGCGCTGTGTACAGCGCCCGCGGAAAGCGTGGCATCTCCGCTGACATC	300						
QY	351	GATCTCTTCCGCGCGAGCGGAGGCCCGCGATCGTCCCATACCGGTTCCAGCGGAAG	410						
DB	301	GATCTCTTCCGCGCGAGCGGAGGCCCGCGATCGTCCCATACCGGTTCCAGCGGAAG	360						
QY	411	AGCACCGTGACCAACCTGCTGGCGGAAATGGCGTGGCGCGGACAAAGGTGTGCGCGTC	470						
DB	361	AGCACCGTGACCAACCTGCTGGCGGAAATGGCGTGGCGCGGACAAAGGTGTGCGCGTC	420						
QY	471	GGCGGCAACTCGGCAACCCCGCGCTGACCTGCTGGCGGACGACATCGAGCTGTACGTG	530						
DB	421	GGCGGCAACTCGGCAACCCCGCGCTGACCTGCTGGCGGACGACATCGAGCTGTACGTG	480						
QY	531	TTGGAGCTCTGAGCTTCAGCTGGAACTCGGATCGGATCGGCTCAACCGCGAGTGGCGAC	590						
DB	481	TTGGAGCTCTGAGCTTCAGCTGGAACTCGGATCGGATCGGCTCAACCGCGAGTGGCGAC	540						
QY	591	GTGCTGAAGCTGAGGCAACCATATGATCGCTACGCGGATGGCTGACTACCACTG	650						
DB	541	GTGCTGAAGCTGAGGCAACCATATGATCGCTACGCGGATGGCTGACTACCACTG	600						
QY	651	GCAAGCACCGATCTTCGCGGTCGCCCGAGGTGCTGATCGCGCGCGATGCCCTG	710						
DB	601	GCAAGCACCGATCTTCGCGGTCGCCCGAGGTGCTGATCGCGCGCGATGCCCTG	660						
QY	711	ACCCGACCGCTGATCGCCGATACCGTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	770						
DB	661	ACCCGACCGCTGATCGCCGATACCGTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	720						
QY	771	TTCAAGGCTTTTCCGCGCTGATGAGGAAGACGGCCAGAGTGGCTGGCGTTCCAGTTCCGAC	830						
DB	721	TTCAAGGCTTTTCCGCGCTGATGAGGAAGACGGCCAGAGTGGCTGGCGTTCCAGTTCCGAC	780						
QY	831	AGCTGCTCGCGGTGGCGAACTGAAGATCCGTGGCGGCCCACTATTTCCAAACCGGTC	890						
DB	781	AGCTGCTCGCGGTGGCGAACTGAAGATCCGTGGCGGCCCACTATTTCCAAACCGGTC	840						
QY	891	GCGCGCTGCGCTGGCGCATCGGTGGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTG	950						
DB	841	GCGCGCTGCGCTGGCGCATCGGTGGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTG	900						
QY	951	AAGGCGTTTTCGCGCTTGCTCATCGTCCGAGTGGTACGAGCGGCGGCGGTGAGC	1010						
DB	901	AAGGCGTTTTCGCGCTTGCTCATCGTCCGAGTGGTACGAGCGGCGGCGGTGAGC	960						
QY	1011	TACTACGAGGATTCCAAAGCCCAACAACTGTCGCGCGCGCGCTGCGCGGCGATCGAGGGGCTG	1070						
DB	961	TACTACGAGGATTCCAAAGCCCAACAACTGTCGCGCGCGCGCTGCGCGGCGATCGAGGGGCTG	1020						
QY	1071	GGTCCGACATCGACGCAAGCTGGTGTGCTGCGCGCGGCGGAGACGCGCAAGGCGCGCAT	1130						
DB	1021	GGTCCGACATCGACGCAAGCTGGTGTGCTGCGCGCGGCGGAGACGCGCAAGGCGCGCAT	1080						
QY	1131	TTCCATGACCTGCGCGAGCGGCTGCGCGCTTCTGCGCGGCGGCTGCTGCTGCTGCTGCTG	1190						
DB	1081	TTCCATGACCTGCGCGAGCGGCTGCGCGCTTCTGCGCGGCGGCTGCTGCTGCTGCTGCTG	1140						
QY	1191	GAGCGCGGCTGATTCGCCAGGCACTGGGCAACGCGGTACCGCTGCTGCTGCTGCTGCTG	1250						
DB	1141	GAGCGCGGCTGATTCGCCAGGCACTGGGCAACGCGGTACCGCTGCTGCTGCTGCTGCTG	1200						
QY	1251	CTGGAGCAAGCACTCGCGAGCGCCCGGAGCTGGCCCGGAGGCGATCGGCTGCTGCTG	1310						
DB	1201	CTGGAGCAAGCACTCGCGAGCGCCCGGAGCTGGCCCGGAGGCGATCGGCTGCTGCTG	1260						
QY	1311	TGCGCGGCTGCGCGAGCTGGACATGTTCAAGAACTTCGAAGAACGCGGAGCGCTGCTG	1370						
DB	1261	TGCGCGGCTGCGCGAGCTGGACATGTTCAAGAACTTCGAAGAACGCGGAGCGCTGCTG	1320						
QY	1371	GCCAAAGCGTAGAGGAGCTAGCGTGA	1397						
DB	1321	GCCAAAGCGTAGAGGAGCTAGCGTGA	1347						
RESULT 2									
US-09-815-242-4132									
; Sequence 4132, Application US/09815242									
; Patent No. US20020061569A1									
GENERAL INFORMATION:									
; APPLICANT: Haselbeck, Robert									
; APPLICANT: Ohlsen, Karl L.									
; APPLICANT: Zyskind, Judith W.									
; APPLICANT: Wall, Daniel									
; APPLICANT: Trawick, John D.									
; APPLICANT: Cart, Grant J.									
; APPLICANT: Yamamoto, Robert T.									
; APPLICANT: Xu, H. Howard									
; TITLE OF INVENTION: Identification of Essential Genes in									
; FILE REFERENCE: ELITRA.011a									
; CURRENT FILING DATE: 2001-03-21									
; PRIOR FILING DATE: 2001-03-21									
; PRIOR APPLICATION NUMBER: 60/191,078									
; PRIOR FILING DATE: 2000-03-21									
; PRIOR APPLICATION NUMBER: 60/206,848									
; PRIOR FILING DATE: 2000-05-23									
; PRIOR APPLICATION NUMBER: 60/207,727									
; PRIOR FILING DATE: 2000-05-26									
; PRIOR APPLICATION NUMBER: 60/242,578									
; PRIOR FILING DATE: 2000-10-23									
; PRIOR APPLICATION NUMBER: 60/253,625									
; PRIOR FILING DATE: 2000-11-27									
; PRIOR APPLICATION NUMBER: 60/257,931									
; PRIOR FILING DATE: 2000-12-22									
; PRIOR APPLICATION NUMBER: 60/269,308									
; PRIOR FILING DATE: 2001-02-16									
; NUMBER OF SEQ ID NOS: 14110									
; SOFTWARE: FastSeq for Windows Version 4.0									
; SEQ ID NO 4132									
; LENGTH: 1200									
; TYPE: DNA									
; ORGANISM: Pseudomonas aeruginosa									
US-09-815-242-4132									
Query Match 3.7%; Score 54; DB 9; Length 1200;									
Best Local Similarity 100.0%; Pred. No. 1.9e-17;									
Matches 54; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
QY	1397	ATGCTGTGCGGTTCGCGCCCTTCCCGCTGCGCGCTGTTGAGCGCGGCGCATC	1450						

Db 1 ATGCTGTCGGTGTTCGCCCTTCCCGTCCGCGCTGTGTGAGCGGCACGGCATC 54

RESULT 3

US-09-815-242-1313/c  
; Sequence 1313, Application US/09815242  
; Patent No. US20020061569A1

GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; TITLE OF INVENTION: Prokaryotes  
; FILE REFERENCE: ELITRA.011A  
; CURRENT FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: US/09/815,242  
; PRIOR FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; NUMBER OF SEQ ID NOS: 14110

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 1313

; LENGTH: 116

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-815-242-1313

Query Match 2.5%; Score 36; DB 9; Length 116;

Best Local Similarity 100.0%; Pred. No. 4.1e-08;

Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1415 CCCTTCCCGTCGCCCTTGTGAGCGGCACGGCATC 1450

Db 116 CCCTTCCCGTCGCCCTTGTGAGCGGCACGGCATC 81

RESULT 4

US-09-815-242-7913  
; Sequence 7913, Application US/09815242  
; Patent No. US20020061569A1

GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; TITLE OF INVENTION: Prokaryotes  
; FILE REFERENCE: ELITRA.011A  
; CURRENT FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: US/09/815,242

; PRIOR FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Kari L.

; APPLICANT: Zyskind, Judith W.

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John D.

; APPLICANT: Carr, Grant J.

; APPLICANT: Yamamoto, Robert T.

; APPLICANT: Xu, H. Howard

; TITLE OF INVENTION: Identification of Essential Genes in

; TITLE OF INVENTION: Prokaryotes

; FILE REFERENCE: ELITRA.011A

; CURRENT FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: US/09/815,242

; PRIOR FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 7913  
; LENGTH: 1083  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)...(1083)  
US-09-815-242-7913

Query Match 2.5%; Score 36; DB 9; Length 1083;

Best Local Similarity 100.0%; Pred. No. 3.1e-08;

Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GTGCTGATCGCGCTCGCCACCTTGAAGCTGCGTTGA 37

Db 1048 GTGCTGATCGCGCTCGCCACCTTGAAGCTGCGTTGA 1083

RESULT 5

US-09-741-669-259

; Sequence 259, Application US/09741669

; Patent No. US20020022718A1

; GENERAL INFORMATION:

; APPLICANT: Forsyth, R. Allyn

; APPLICANT: Ohlsen, Kari L.

; APPLICANT: Zyskind, Judith W.

; TITLE OF INVENTION: Genes identified as required for

; FILE REFERENCE: ELITRA.009A

; CURRENT FILING DATE: 2000-12-19

; PRIOR APPLICATION NUMBER: US 60/173005

; PRIOR FILING DATE: 1999-12-23

; NUMBER OF SEQ ID NOS: 481

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 259

; LENGTH: 1317

; TYPE: DNA

; ORGANISM: Escherichia coli

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (1)...(1317)

US-09-741-669-259

Query Match 1.7%; Score 24; DB 9; Length 1317;

Best Local Similarity 100.0%; Pred. No. 0.041;

Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 537 CTGTCGAGCTTCCAGCTGGAACCC 560

Db 475 CTGTCGAGCTTCCAGCTGGAACCC 498

RESULT 6

US-09-815-242-5932

; Sequence 5932, Application US/09815242

; Patent No. US20020061569A1

; GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Kari L.

; APPLICANT: Zyskind, Judith W.

; TITLE OF INVENTION: Identification of Essential Genes in

; TITLE OF INVENTION: Prokaryotes

; FILE REFERENCE: ELITRA.011A

; CURRENT FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: US/09/815,242

; PRIOR FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Kari L.

; APPLICANT: Zyskind, Judith W.

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John D.

; APPLICANT: Carr, Grant J.

; APPLICANT: Yamamoto, Robert T.

; APPLICANT: Xu, H. Howard

; TITLE OF INVENTION: Identification of Essential Genes in

; TITLE OF INVENTION: Prokaryotes

; FILE REFERENCE: ELITRA.011A

; CURRENT FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: US/09/815,242

; PRIOR FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Kari L.

; APPLICANT: Zyskind, Judith W.

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John D.

; APPLICANT: Carr, Grant J.

; APPLICANT: Yamamoto, Robert T.

; APPLICANT: Xu, H. Howard

; TITLE OF INVENTION: Identification of Essential Genes in

; TITLE OF INVENTION: Prokaryotes

; FILE REFERENCE: ELITRA.011A

; CURRENT FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: US/09/815,242

; PRIOR FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

20

Db 2507 GCAGGCGCGGAGCTGGCGC 2488

## RESULT 10

US-09-927-827-1/c  
; Sequence 1, Application US/09927827  
; Publication No. US20030036176A1  
; GENERAL INFORMATION:  
; APPLICANT: Bower, Stanley G.  
; APPLICANT: Ramseier, Thomas M.  
; TITLE OF INVENTION: Directed Genetic Engineering of Xanthomonas campestris  
; FILE REFERENCE: 38-10(15824)B  
; CURRENT APPLICATION NUMBER: US/09/927,827  
; CURRENT FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: US 60/279,493  
; PRIOR FILING DATE: 2001-03-28  
; NUMBER OF SEQ ID NOS: 69  
; SEQ ID NO 1  
; LENGTH: 7356  
; TYPE: DNA  
; ORGANISM: Xanthomonas campestris  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(7356)  
; OTHER INFORMATION: unsure at all n locations  
US-09-927-827-1

Query Match 1.4%; Score 20; DB 11; Length 7356;

Best Local Similarity 100.0%; Pred. No. 3.6;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 500 CCGTGGCGGAGCATCG 519

|||||

Db 6090 CCGTGGCGGAGCATCG 6071

## RESULT 11

US-10-156-761-1/c  
; Sequence 1, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 1  
; LENGTH: 9025608  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: misc.feature  
; LOCATION: (4187715)  
; OTHER INFORMATION: a, t, c, g, other or unknown  
US-10-156-761-1

Query Match 1.4%; Score 20; DB 14; Length 9025608;

Best Local Similarity 100.0%; Pred. No. 1.4;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1094 GGTGCTGCTCGCGCGGAG 1113

|||||

Db 6407532 GGTGCTGCTCGCGCGGAG 6407513

## RESULT 12

US-10-156-761-3485  
; Sequence 3485, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 3485  
; LENGTH: 393  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(393)  
US-10-156-761-3485

Query Match 1.3%; Score 19; DB 14; Length 393;

Best Local Similarity 100.0%; Pred. No. 17;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1270 AGCGCGCGGAGTGGCCG 1288

|||||

Db 143 AGCGCGCGGAGTGGCCG 161

## RESULT 13

US-10-084-843-268  
; Sequence 268, Application US/10084843  
; Publication No. US20030143243A1  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Dillon, Davin C.  
; APPLICANT: Campos-Neto, Antonio  
; APPLICANT: Houghton, Raymond  
; APPLICANT: Vedvick, Thomas S.  
; APPLICANT: Twardzik, Daniel R.  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Hendrickson, Ronald C.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY  
; FILE REFERENCE: 355  
; CURRENT APPLICATION NUMBER: US/10/084,843  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 1  
; LENGTH: 9025608  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: misc.feature  
; LOCATION: (4187715)  
; OTHER INFORMATION: a, t, c, g, other or unknown  
US-10-084-843-268

CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED and BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/084,843

;; FILING DATE: 25-Feb-2002  
;; CLASSIFICATION: <Unknown>  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US/09/072,967  
;; FILING DATE: 05-MAY-1998  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Makl, David J.  
;; REGISTRATION NUMBER: 31,392  
;; REFERENCE/DOCKET NUMBER: 210121.411C9  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (206) 622-4500  
;; TELEFAX: (206) 682-6031  
;; INFORMATION FOR SEQ ID NO: 268:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 522 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: cDNA  
;; SEQUENCE DESCRIPTION: SEQ ID NO: 268:  
US-10-084-843-268  
  
Query Match 1.3%; Score 19; DB 12; Length 522;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1082 CGACGGCAAGCTGGTCTG 1100  
Db 429 CGACGGCAAGCTGGTCTG 447  
  
RESULT 14  
US-10-193-002-263  
; Sequence 263, Application US/10193002  
; Publication No. US20030135026A1  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; Skeiky, Yasir A.W.  
; Dillon, Davin C.  
; Campos-Neto, Antonia  
; Houghton, Raymond  
; Vedvick, Thomas S.  
; Twardzik, Daniel R.  
; Lodes, Michael J.  
; Hendrickson, Ronald C.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF  
; TUBERCULOSIS  
; NUMBER OF SEQUENCES: 350  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED and BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104-7092  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/10/193,002  
; FILING DATE: 10-Jul-2002  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/09/072,596  
; FILING DATE: 05-MAY-1998  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Makl, David J.  
; REGISTRATION NUMBER: 31,392  
; REFERENCE/DOCKET NUMBER: 210121.417C9  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4500

;; TELEFAX: (206) 682-6031  
;; INFORMATION FOR SEQ ID NO: 263:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 522 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: cDNA  
;; SEQUENCE DESCRIPTION: SEQ ID NO: 263:  
US-10-193-002-263  
  
Query Match 1.3%; Score 19; DB 12; Length 522;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1082 CGACGGCAAGCTGGTCTG 1100  
Db 429 CGACGGCAAGCTGGTCTG 447  
  
RESULT 15  
US-10-156-761-7015  
; Sequence 7015, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 7015  
; LENGTH: 648  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(648)  
US-10-156-761-7015  
  
Query Match 1.3%; Score 19; DB 14; Length 648;  
Best Local Similarity 100.0%; Pred. No. 16;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 CGTGCTGATCGGCTCGCC 19  
Db 546 CGTGCTGATCGGCTCGCC 564  
  
Search completed: August 14, 2003, 07:13:43  
Job time : 437 secs

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run On: August 14, 2003, 08:41:46 ; Search time 4238 seconds  
(without alignments)  
4324.566 Million cell updates/sec

Title: US-09-701-229-2

Perfect score: 448

Sequence: 1 MSLIASDHFRIVVGLKSGM.....MFKNFEERGLFAKAVEELA 448

Scoring table: OLIGO /  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 2045481386 residues

Word size: 1

Total number of hits satisfying chosen parameters: 4990538

Minimum DB seq length: 25

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:  
-MODEL=frame+p2n.model -DEV=xlp  
-O=/cgn2.1/USPTO.spool/US09701229/runat\_09082003\_161236\_5021/app.query.fasta.1.647  
-DB=GenEmbl -QFMT=fastap -SUFFIX=olip2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=Dits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45  
-DOCCALIGN=200 -THR\_SCORE=quality -THR\_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTPMT=pto  
-NORM=ext -HEAPSIZ=500 -MINLEN=25 -MAXLEN=2000000000  
-USER=US09701229.@CGN.1.1.5265.@runat\_09082003\_161236\_5021 -NCPU=3  
-NO\_MMAPP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

GenEmbl.\*

1: gb\_ba.\*  
2: gb\_htg.\*  
3: gb\_in.\*  
4: gb\_ov.\*  
5: gb\_pa.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vi.\*  
15: em\_ba.\*  
16: em\_fun.\*  
17: em\_hum.\*  
18: em\_in.\*  
19: em\_lm.\*  
20: em\_om.\*  
21: em\_or.\*  
22: em\_ov.\*  
23: em\_ph.\*  
24: em\_pl.\*  
25: em\_ro.\*  
26: em\_ro.\*  
27: em\_un.\*  
28: em\_un.\*

29: em\_vi.\*  
30: em\_htg\_hum.\*  
31: em\_htg\_inv.\*  
32: em\_htg\_other.\*  
33: em\_htg\_mus.\*  
34: em\_htg\_pln.\*  
35: em\_htg\_rod.\*  
36: em\_htg\_mam.\*  
37: em\_htg\_vrt.\*  
38: em\_sy.\*  
39: em\_htgo\_hum.\*  
40: em\_htgo\_mus.\*  
41: em\_htgo\_other.\*

pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
c 1	448	100.0	24000	1	AE004856 Pseudomon
2	290	64.7	5280	1	AY008276 Pseudomon
3	35	7.8	301995	1	AE016779 Pseudomon
c 4	20	4.5	311600	1	AE016871 Pseudomon
c 5	19	4.2	96109	6	AX067460 Sequence
6	14	3.1	1317	6	AX189058 Sequence
7	14	3.1	1629	1	ECMURD
8	14	3.1	2608	1	ECMUROY
9	14	3.1	11441	1	AE015046
10	14	3.1	11498	6	U32793
11	14	3.1	11498	6	AX191762 Sequence
12	14	3.1	12791	1	AE005185 Escherich
13	14	3.1	21757	1	AE000118 Escherich
c 14	14	3.1	22201	1	AE004310
15	14	3.1	22348	1	AE008699
16	14	3.1	28277	6	AX191720
17	14	3.1	28277	15	EC2MIN
18	14	3.1	110000	6	AR274513_11
19	14	3.1	110000	6	AR274513_12
20	14	3.1	111408	1	EC0110K
21	14	3.1	248308	2	AC073937
22	14	3.1	251050	1	AL627265
c 23	14	3.1	259498	2	AC020876
24	14	3.1	281530	1	AP002550
25	14	3.1	290029	1	AE016978
26	14	3.1	300169	1	AE016834
27	14	3.1	300409	1	AE016755
c 28	14	3.1	301442	1	AE016798
29	13	2.9	597	1	AY142808
c 30	13	2.9	14643	1	AE014097
31	13	2.9	298900	1	AP005074
32	13	2.9	316050	1	BX321859
c 33	12	2.7	11726	1	AE013965
34	12	2.7	12232	1	AE006048
c 35	12	2.7	13071	1	AE015855
36	12	2.7	21000	1	AB052554
c 37	12	2.7	188050	1	AL646072
38	12	2.7	201050	1	AJ414143
39	12	2.7	301727	1	AE016960
c 40	12	2.7	348600	1	AB063521
41	11	2.5	1497	1	EC0FTSW
42	11	2.5	110000	2	LMFLCHR34_08
c 43	11	2.5	166560	5	AL954741
44	10	2.2	10629	1	AE002398
c 45	10	2.2	11683	1	AE013119

#### ALIGNMENTS

RESULT 1

```

AE004856/C
LOCUS
DEFINITION
  AE004856      24000 bp      DNA      linear      BCT 19-FEB-2003
  Pseudomonas aeruginosa PAO1, section 417 of 529 of the complete
  genome.
ACCESSION
  AE004856      AE004091
VERSION
  AE004856.1    GI:9950633
SOURCE
  Pseudomonas aeruginosa PAO1
ORGANISM
  Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
  Pseudomonadaceae; Pseudomonas.
REFERENCE
  1 (bases 1 to 24000)
  Stover, C.K., Pham, X.-Q.T., Erwin, A.L., Mizoguchi, S.D., Warren, P.,
  Hickey, M.J., Brinkman, F.S.L., Hufnagle, W.O., Kowalik, D.J.,
  Lagrou, M., Garber, R.L., Goltry, L., Tolentino, E.,
  Westbrook-Wadman, S., Yuan, Y., Brody, L.L., Coulter, S.N.,
  Folger, K.R., Kas, A., Larbig, K., Lim, R.M., Smith, K.A., Spencer, D.H.,
  Wong, G.K.-S., Wu, Z., and Paulsen, I.T.
  Complete genome sequence of Pseudomonas aeruginosa PAO1, an
  opportunistic pathogen
  Nature 406 (6799), 959-964 (2000)
  20437337
  10984043
REFERENCE
  2 (bases 1 to 24000)
  Stover, C.K., Pham, X.-Q.T., Erwin, A.L., Mizoguchi, S.D., Warren, P.,
  Hickey, M.J., Brinkman, F.S.L., Hufnagle, W.O., Kowalik, D.J.,
  Lagrou, M., Garber, R.L., Goltry, L., Tolentino, E.,
  Westbrook-Wadman, S., Yuan, Y., Brody, L.L., Coulter, S.N.,
  Folger, K.R., Kas, A., Larbig, K., Lim, R.M., Smith, K.A., Spencer, D.H.,
  Wong, G.K.-S., Wu, Z., Paulsen, I.T., Reizer, J., Saier, M.H.,
  Hancock, R.E.W., Lory, S. and Olson, M.V.
  Direct Submission
  Submitted (16-MAY-2000) Department of Medicine and Genetics,
  University of Washington Genome Center, University Of Washington,
  Box 352145, Seattle, WA 98195, USA
  3 (bases 1 to 24000)
  Pseudomonas aeruginosa Community Annotation Project (PseudoCAP)
  Direct Submission
  Submitted (04-FEB-2003) Department of Molecular Biology and
  Biochemistry, Simon Fraser University, 8888 University Dr.,
  Burnaby, British Columbia V5A 1S6, Canada
  -----
  This represents the February 3, 2003 version of the continually
  updated, reviewed, Pseudomonas aeruginosa PAO1 genome annotation,
  from PseudoCAP (see http://www.pseudomonas.com for latest updates
  and links to alternate annotations). PseudoCAP is coordinated by
  Fiona S.L. Brinkman (Simon Fraser University, Canada) and Robert
  E.W. Hancock (University of British Columbia, Canada). We welcome
  submission through www.pseudomonas.com of any proposed changes.
  -----
  'Protein name confidence' is used to rate our confidence of the
  accuracy of the protein name.
  Class 1: Function experimentally demonstrated in P. aeruginosa.
  Class 2: Function of highly similar gene experimentally
  demonstrated in another organism (and gene context consistent
  in terms of pathways its involved in, if known).
  Class 3: Function proposed based on presence of conserved amino
  acid motif, structural feature or limited sequence similarity
  to an experimentally studied gene.
  Class 4: Homologs of previously reported genes of unknown function,
  or no similarity to any previously reported sequences.
  -----
  FEATURES
  source
    Location/Qualifiers
      1..24000
        /organism="Pseudomonas aeruginosa PAO1"
        /mol_type="genomic DNA"
        /strain="PAO1"
        /db_xref="taxon:208964"
        complement(91..1308)
        /gene="argJ"
        /locus_tag="PA4402"
        complement(91..1308)
  gene
  CDS
  /gene="argJ"
  /locus_tag="PA4402"
  /note="Protein name confidence: Class 2 (high similarity
  to functionally studied protein); Subcellular
  localization: Cytoplasmic; Subcellular localization
  confidence: Class 2; Alt protein name: ornithine
  transacetylase"
  /codon_start=1
  /transl_table=11
  /product="glutamate N-acetyltransferase"
  /protein_id="AA07790.1"
  /db_xref="GI:9950634"
  /translation="MAVGGLPLSTLHPVPELGIASAGIKRGRKVVVMCAEAGSS
  VAGVFTLNAFCAPVTAKQPLGEVRYLLTNTGNANCTGAGLAAQTCAKLAEL
  AGVATSVLPYGTGVEGPLVPAKIEAALPAALADLAEDRWAEAAAGITWDTLPKGA
  SQFVHDGVTVVTGIGSKAGMKPNMATMLGVYATDAKVAQVQLDLDRDAANKSFN
  RITIDGTSNDCCLLIATGRAALPEVTOASGALPAALQAQVLSVSMELAQAIVRGE
  GATKFTVGVNGATHQECIDVGYAVAHSPLIKLTALFASDPNWRGLAAVGRGVANL
  DYSKIDVFLGDVCIASRGRASYTEEQGAAYMAQAEIGIRIELGRGTCSETIWTDTL
  SHEYVKINAEYRT"
  complement(1455..4205)
  /gene="secA"
  /locus_tag="PA4403"
  complement(1455..4205)
  /gene="secA"
  /locus_tag="PA4403"
  /note="Protein name confidence: Class 2 (high similarity
  to functionally studied protein)"
  /codon_start=1
  /transl_table=11
  /product="secretion protein SecA"
  /protein_id="AA07791.1"
  /db_xref="GI:9950635"
  /translation="MFAPLKKLFGSKNERDVKRMKAVQATNALEPQMVALSDEQLK
  AKTAEQQRYAKGETLDQLLPFAVAVREAGKRVGMGRHFDYQLIGGWTLHGKTAEM
  RTGKTLVGTLPVYNALSGVGVVTVNDLARDNANMRPLREFLGLSVGVVTFP
  QPEDKRAAYAAADITYGNNEFDYLNMAFSLDDKQFRELNFVAVDEVDSILIDE
  ARTPLIISQAEDSSELYKINKPLRKQVEVEGKTEEGHYSIDKTRQVLENE
  OCHQFTIEDLLSQNGELGEGYSAHNLSLLTHVYAAALRAHTFTHRNVEYIVQGOQIL
  LIDEHTGTRPGRRLSEGLHQAIEAKGLPIQAESOTLASTTFQNFVRYNKLAGWTG
  TADTEAFERFOIYGLDVVPIPHRTIARKDFNDLYLTQEEKYAITIDIKCOALGR
  PILVGTASIESSEYVSKLQEAGIEHKVLNAYKEAEIIIAQAGAPSVTATNMG
  RGTDLILGNGNEVEVALENPTTEQIAKAEWKRRHQVIEAGGLHVTSERHESRR
  IDNQRAGRGQDGPSSRFYLSLEDNLMRIFASDRVKNFMKALMQSGEATEHRMVT
  NAIKQARKVEGRNFDIRKQLDFDVAHQKRVYTHMRNTLLSADDEGTKEFREEL
  TLSATINOHIPPOSUPEQDIEGLEAALYSDFAVRLPTQOWLDEDKLYEETLRKIL
  FOIVAAVYEKEELAGALRAPEKOMLLRVLDDLAKDHLSTMDHRLHGTHLGRYAKQN
  PQEYKRSFTLQELLSIRKDTIRVLSHVQVREDPAEEARLRREAEELAKRMQF
  QHAEAPSEQAVAGEEELPFGVPVPLEPVVRNEQKIGRNEPCPCGSGKTKHCHGQ
  LD"
  complement(4339..5262)
  /locus_tag="PA4404"
  complement(4339..5262)
  /locus_tag="PA4404"
  /note="Protein name confidence: Class 4 (homologs of
  previously reported genes of unknown function, or no
  similarity to any previously reported sequences)"
  /codon_start=1
  /transl_table=11
  /product="hypothetical protein"
  /protein_id="AA07792.1"
  /db_xref="GI:9950636"
  /translation="MHIIFLSRHHGAARSVHLOPRLLLAGACVYVGSLLAGGFMGAW
  LFRAPAVAGIEVDAAALQAQSAEISSAREDAQOMDALSVLHGLQALRLVRLDAGER
  LTELADVPDIEFDFSLSYOGGVDEPLEGSVAPPSPFATLDELARLDTREQOOLEVL
  EQLLADKHISOSTYLAGOPVRQGYTSSPFGRRVNPVSGRLSHHGKIDIFAAPAGSDVVA
  VAGVYVHSGRKNGYGVNVEVGHADGYTTLTAHNQKNLYVEVGDLYVKRQVLAQVSTG
  RSTGYVHVFVNMKDRGNVNPQSFTARASVSE"
  5409..5804
  /locus_tag="PA4405"
  5409..5804
  /locus_tag="PA4405"
  /note="Protein name confidence: Class 4 (homologs of

```



previously reported genes of unknown function, or no similarity to any previously reported sequences)

```
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA007793.1"
/db_xref="GI:9950637"
/translation="MLFAEQRICRLQOLLESQIOPAAAPYCHVASWRGCLLIVTD
GHVATRLRYOORLORLOQAEFEFANAKILFKVQPTTSQRAARRSLTISAAEENL
RESAEIVGNPRLREALERLASRAKRSQE"
/complement(5865..6776)
/gene="lpxc"
```

gene

```
/locus_tag="PA4406"
/notes="synonyms: envA, asmB"
/complement(5865..6776)
/gene="lpxc"
/locus_tag="PA4406"
```

CDS

```
/note="Protein name confidence: Class 1 (function
experimentally demonstrated in P. aeruginosa); Alt protein
name: EnvA protein"
/codon_start=1
```

```
/transl_table=11
/product="UDP-3-O-acetyl-N-acetylglucosamine deacetylase"
/protein_id="AA007794.1"
/db_xref="GI:9950638"
```

```
/translation="MIKQRTLNIRATGVGLHSGEKVYLTLPAPVDTGIVFCRTDL
DPVVEIPARAENVGETTMTSTLVKGDVVDTHLLSAMAGLIGDINAYVELSASEVPI
MDGSAGPVELIQSAGLOEQAAKFIRIKREVSVEEDKRAVFPVFDGKVSPEIDF
DPVFRGRTQOASVDFSTSFVKEVSARTFGEMRDIEYLRSONLAGGVSVENAIVVD
ENRVNEDGRLREDFVKHILKIDGLYLLGNSLIGEFGRFSGHALNNQLTLIA
DKDAWEVTFEDARTAPISYMPAAAV"
```

gene

```
/gene="ftsZ"
/locus_tag="PA4407"
/complement(6889..8073)
```

CDS

```
/gene="ftsZ"
/locus_tag="PA4407"
/complement(6889..8073)
```

```
/note="Protein name confidence: Class 2 (high similarity
to functionally studied protein)"
/codon_start=1
```

```
/transl_table=11
/product="cell division protein FtsZ"
/protein_id="AA007795.1"
/db_xref="GI:9950639"
/translation="MFELVDNIAQTAVIKVIGVGGGNNVNHMAKNVGEFICAN
TDAALANIARTVLQPGVTGKLGAGANPEVGRQAALEDREIRISVLEGADWVFT
TGNGSGTGAAPILAEVAKEMGILLVAVTRPPFPFGRKRMQIADSGIRALAESVDS
LITPNEKLLTIGKASLLAFAKADVDVLAGAVRGSIDIKRPGMINVDFADVKTVM
SEMGMWMTGCASGPNRAREATEAARINPLLEDVNIQAGRGILVNTAGPDVLSLEY
SDVGNIIEQFASERATVKGTVIDMRDELHVTVAATGLGARLEKPKVKVYDNTVQGS
AAQAAAPAREQSVNVRDLRPTVMRNQSHGSAATAAKLNPDLDLDIPAFLRQ
AD"
/complement(8124..9377)
/gene="ftsA"
```

gene

```
/locus_tag="PA4408"
```

## Alignment Scores:

Pred. No.:	Score:	Length:	Matches:
0	448.00	24000	448
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	1	Gaps:	0

US-09-701-229-2 (1-448) x AE004856 (1-24000)

Qy	1	MetSerLeuIleAlaSerAspHisPheArgIleValValGlyLeuGlyLysSerGlyMet	20
Db	16265	ATGAGCTGATCGCTCGACCACTTCGGCATCTTGTGGCCTCGGCAAGCGGCATG	16206
Qy	21	SerLeuValArgTyrLeuAlaArgGlyLeuProPheAlaValAlaValAspThrArgGlu	40
Db	16205	TCCCTGTGCTACCTGGCGCGCGGCTTGCCTTTCGCGGTGGTTCGATACCCGAGAG	16146

Qy	41	AsnProGluLeuAlaThrLeuArgAlaGlnTyrProGlnValGluValArgCysGly	60
Db	16145	AACCCGCGGAGCTGGCCACCCTCGGTCGCCAGTATCCGCAGGTGGAAAGTCCGTTGCCGC	16086
Qy	61	GluLeuAspAlaGluPheLeuCysSerAlaArgGluLeuTyrValSerProGlyLeuSer	80
Db	16085	GAATCGACGCGGAGTTCTCTGCTCGCCCGCGCAACTCTATGTTCAGCCCGGCTTTCG	16026
Qy	81	LeuArgThrProAlaLeuValGlnAlaAlaLysGlyValArgIleSerGlyAspIle	100
Db	16025	CTGCGCACCCCTGCGCTGTGTACAGCGCGCCGAAAGCGTGCATCTCCGGTGACATC	15966
Qy	101	AspLeuPheAlaArgGluAlaLysAlaProIleValAlaIleThrGlySerAsnAlaLys	120
Db	15965	GATCTCTTCGCGCGGAGCGAAGCCCGCATCTCGGCATCACCGGTTCACACCGGAAG	15906
Qy	121	SerThrValThrLeuValGlyGluMetAlaValAlaAlaAspLysArgValAlaVal	140
Db	15905	AGCACCGTGCACACCTGGTGGGGAATGGCGGTGGCGCGGACCAAGCGTGTCCCGCTC	15846
Qy	141	GlyGlyAsnLeuGlyThrProAlaLeuAspLeuAlaAspAspIleGluLeuTyrVal	160
Db	15845	GCGCGCAACCTCGGCACCCCGCGCTCGACCTGCTGGCGCAGCATCGAGCTGACGTG	15786
Qy	161	LeuGluLeuSerSerPheGlnLeuGluThrCysAspArgLeuAsnAlaGluValAlaThr	180
Db	15785	TTGGAGCTGTGAGCTTCCAGCTGGAACCTGCGATCGCTCAACGCCGAGGTGCGGACC	15726
Qy	181	ValLeuAsnValSerGluAspHisMetAspArgTyrAspGlyMetAlaAspTyrHisLeu	200
Db	15725	GTGCTGACGTCAGCAGACCATATGATGCTGCTACGACGCGCATGGCTGACTACCACTG	15666
Qy	201	AlaLysHisArgIlePheArgGlyAlaArgGlnValValAlaAspArgAlaAspAlaLeu	220
Db	15665	GCCAAGCACCGGATCTTCCGCGGTGCCCGCAGGTGCTGGTGAATTCGCCCGCATGCCCTG	15606
Qy	221	ThrArgProLeuIleAlaAspThrValProCysTrpSerPheGlyLeuAsnLysProAsp	240
Db	15605	ACCCGACGCTGATCGCGCATACCGTGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	15546
Qy	241	PhelysAlaPheGlyLeuIleGluAspGlyGlnLysTrpLeuAlaPheGlnPheAsp	260
Db	15545	TTCAAGGCTTTCGGCTGATCGAGGAGACGCCAGAGTGGCTGGCTGCTGCTGCTGCTGCTG	15486
Qy	261	LysLeuLeuProValGlyLeuLysIleArgGlyAlaHisAspTyrSerAsnAlaLeu	280
Db	15485	AAGCTGCTGCGGTGGCGAACTGAAGATCGTGGCGCCCACTATTCACACCGCTC	15426
Qy	281	AlaAlaLeuAlaLeuGlyHisAlaValGlyLeuProPheAspAlaMetLeuGlyAlaLeu	300
Db	15425	GCCGCGCTGGCGCTGGCGCATCGGTCGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTG	15366
Qy	301	LysAlaPheSerGlyLeuAlaHisArgCysGlnTrpValArgGluArgGlyValSer	320
Db	15365	AAGCGTTCCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	15306
Qy	321	TyrTyrAspAspSerLysAlaThrAsnValGlyAlaAlaLeuAlaAlaIleGluLeu	340
Db	15305	TACTACGACGATTCAGGCCACCAACCTGCGCGCGCTGCGCGCGCTGCGCGCGCTGCGCGCG	15246
Qy	341	GlyAlaAspIleAspGlyLysLeuValLeuLeuAlaGlyGlyAspGlyLysGlyAlaAsp	360
Db	15245	GCTGCCGACATCGACGCGAAGCTGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	15186
Qy	361	PheHisAspLeuArgGluProValAlaArgPheCysArgAlaValAlaLeuLeuGlyArg	380
Db	15185	TTCCATGACCTGCGAGCGGCTCGCGGCTTTCGCGCGGCTGCTGCTGCTGCTGCTGCTGCTG	15126
Qy	381	AspAlaGlyLeuIleAlaGlnAlaLeuGlyAsnAlaValProLeuValArgValAlaThr	400
Db	15125	GACCGCGGCTGATTGCCAGGCACTGGGCAACCGGTCACCGCTGCTGCTGCTGCTGCTGCTG	15066



```

Db 3929 ATGAGCCTGATCCCTCCGACCACTTCCCGCATCGTTGTGCGCTCCGGCAAGACGGCGATG 3988
QY 21 SerLeuValArgTyrLeuAlaArgArgGlyLeuProPheAlaValValAlaAspThrArgGlu 40
Db 3989 TCCTGTGTGGCTACCTTGGCGCGCGCGGCTTGCCTTTCCGCTGTCGATACCCGAGAG 4048
QY 41 AsnProProGluLeuAlaThrLeuArgAlaGlnTyrProGlnValGluVal -ArgCysGI 60
Db 4049 AACCCCGCGAGGTGGCTACCTTGGCTGCGTCCGACATATCCGAGGTGGAAGC -GGGTTGCGG 4107
QY 60 yGluLeuAspAlaGluPheLeuCysSerAlaArgGluLeuTyrValSerProGlyLeuSe 80
Db 4108 CGAACTCGACCGCGAGTTCTCTGCTCCGCCCGCGAACTCTATGTCAGCCCGCGCTGTC 4167
QY 80 rLeuArgThrProAlaLeuValGlnAlaAlaLysGlyValArgLysSerGlyAspII 100
Db 4168 GCTGCGCACCCCTGCGCTGGTACAGCGCGCGGAAAGGCGTGCATCTCCGCTGCAT 4227
QY 100 eAspLeuPheAlaArgGluAlaLysAlaProIleValAlaIleThrGlySerAsnAlaLy 120
Db 4228 CGATCTCTTGGCCCGCGAGCGAAGCCCGGATCGTCCGCTACCTCCAGCGTTCCAACGCGAA 4287
QY 120 sSerThrValThrLeuValGlyGluMetAlaValAlaAlaAspLysArgValAlaVa 140
Db 4288 GACACCGGTGACCACTTGGTGGCGGAAATGGCGGTGGCGCGGACAAGCGTGTCCGCT 4347
QY 140 lGlyLysAsnLeuGlyThrProAlaLeuAspLeuLeuAlaAlaAspIleGluLeuTyrVa 160
Db 4348 CGCGCGCAACCTCGCGCACCCCGCGCTGACCTGCTGGCGGACGACATCGAGCTGTACGT 4407
QY 160 lLeuGluLeuSerSerPheGlnLeuGluThrCysAspArgLeuAsnAlaGluValAlaTh 180
Db 4408 GTTGGAGCTGTCCAGCTTCCAGCTGGAAACCTCGATCGCTCAACGCCGCGAGTGGCGAC 4467
QY 180 rValLeuAsnValSerGluAspHisMetAspArgTyrAspGlyMetAlaAspTyrHisLe 200
Db 4468 CGTCTGAACTGAGCGTGGAGACCATATGATGCTGCTAGCGGATGGCTGACTACCACT 4527
QY 200 uAlaLysHisArgIlePheArgGlyAlaArgGlnValValValAsnArgAlaAspAlaLe 220
Db 4528 GGCCAAAGCACCGGATCTTCGCGGTGCCCGCAGGTGCTGGTGAATCGCGCGCATGCCCT 4587
QY 220 uThrArgProLeuIleAlaAspThrValProCysTyrTrpSerPheGlyLeuAsnLysProAs 240
Db 4588 GACCCGACGCTGATCGCGATACCGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4647
QY 240 pPheLysAlaPheGlyLeuIleGluAspGlyGlnLysTrpLeuAlaPheGlnPheAs 260
Db 4648 CTTCAAGGCTTTCGGCTGATCGAGGAAGACGGCCAGAGTGGCTGGCTGCTGCTGCTGCTG 4707
QY 260 p-LysLeuLeuProValGlyGluLeuLysIleArgGlyAlaHisAsnTyrSerAsnAlaL 280
Db 4708 -CAAGCTGCTGGGTTGGGAACTCAAGATCGTGGCGCCCACTATTCACGCGCG 4766
QY 280 euAlaAlaLeuAlaLeuGlyHisAlaValAlaGlyLeuProPheAspAlaMetLeuGlyAlaL 300
Db 4767 TCGCCGCGTGGCGTGGCGCATCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCG 4826
QY 300 euLysAlaPheSerGlyLeuAlaHisArgCysGlnTrpValArgGluArgGlnGlyValS 320
Db 4827 TGAAGCGGTTTCGGCGCTGGCTCATCGTGGCAGTGGGTACGCGAGCGGAGGCGGTGA 4886
QY 320 eTyrTyrAspSerLysAlaThrAsnValGlyAlaAlaLeuAlaAlaIleGluGlyL 340
Db 4887 GCTACTACGAGATTCCAGGCCACCAAGTGGCGCGCGCTGGCGCGGATCGAGGGCG 4946
QY 340 euGlyAlaAspIleAspGlyLysLeuValLeuLeuAlaGlyGlyAspGlyLysGlyAlaA 360
Db 4947 TGGGTGCCACATCGACGCGCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5006
QY 360 spPheHisAspLeuArgGluProValAlaArgPheCysArgAlaValValLeuLeuGlyA 380
Db 5007 ATTTCCATGACCTGCGCGAGCGCGCTTCTTCTGCGGGCGGTGGTACTGCTTGGCC 5066

```

```

QY 380 rGspAlaGlyLeuIleAlaGlnAlaLeuGlyAsnAlaValProLeuValArgValAlaAt 400
Db 5067 GTGACGCGCGGCTGATTGCCAGGCACTGGGCAACGCGGTACCGTGTGGCGCTGCCAA 5126
QY 400 hrLeuaspGluAlaValArgGlnAlaAlaGluLeuAlaArgGluGlyAspAlaValLeuL 420
Db 5127 CGTGGACGAAGCAGTCCGCGAGCCGCGAGTGGCCCGCGAAGCGGATGCGGTGTGT 5186
QY 420 euSerProAlaCysAlaSerLeuAspMetPheLysAsnPheGluGluArgGlyArgLeuP 440
Db 5187 TGTGCGCGGCTGCGCGAGCTTGACATGTTTCAAGAACTTCGAAGAAGACGCGGACGCTGT 5246
QY 440 healLysAlaValGluGluLeuAla 448
Db 5247 TCGCAAAGCCGTAGAGGAGTAGCG 5272

RESULT 3
AE016779
LOCUS
DEFINITION
Pseudomonas putida KT2440 section 6 of 21 of the complete genome.
ACCESSION
AE016779 AE015451
VERSION
AE016779.1 GI:26557023
KEYWORDS
Pseudomonas putida KT2440
SOURCE
Pseudomonas putida KT2440
ORGANISM
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE
1 (bases 1 to 301995)
Nelson,K., Paulsen,I., Weinel,C., Dodson,R., Hilbert,H., Fouts,D.,
Gill,S., Pop,M., Martins Dos Santos,V., Holmes,M., Brinkac,L.,
Beanan,M., DeBoy,R., Daugherty,S., Kolonay,J., Madupu,R.,
Nelson,W., White,O., Peterson,J., Khouri,H., Hance,I., Lee,P.,
Holtzapple,E., Scanlan,D., Tran,K., Moazzzez,A., Utterback,T.,
Rizzo,M., Lee,K., Kosack,D., Moestl,D., Wedler,H., Lauber,J.,
Hoheisel,J., Straetz,M., Helm,S., Kiewitz,C., Eisen,J., Timmis,K.,
Duesterhoft,A., Tummeler,B. and Fraser,C.
Complete genome sequence and comparative analysis of the
metabolically versatile Pseudomonas putida KT2440
Environ. Microbiol. 4 (12), 799-808 (2002)
2 (bases 1 to 301995)
Nelson,K., Paulsen,I., Weinel,C., Dodson,R., Hilbert,H., Fouts,D.,
Gill,S., Pop,M., Martins Dos Santos,V., Holmes,M., Brinkac,L.,
Beanan,M., DeBoy,R., Daugherty,S., Kolonay,J., Madupu,R.,
Nelson,W., White,O., Peterson,J., Khouri,H., Hance,I., Lee,P.,
Holtzapple,E., Scanlan,D., Tran,K., Moazzzez,A., Utterback,T.,
Rizzo,M., Lee,K., Kosack,D., Moestl,D., Wedler,H., Lauber,J.,
Hoheisel,J., Straetz,M., Helm,S., Kiewitz,C., Eisen,J., Timmis,K.,
Duesterhoft,A., Tummeler,B. and Fraser,C.
Direct Submission
Submitted (05-NOV-2002) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
FEATURES
source
1. 301995
/organism="Pseudomonas putida KT2440"
/mol_type="genomic DNA"
/strain="KT2440"
/db_xref="taxon:160488"
89. .712
/gene="sspA"
/locus_tag="PP1320"
89. .712
/gene="sspA"
/locus_tag="PP1320"
/note="similar to GP:9864180; identified by sequence
similarity; putative"
/transl_table=11
/product="stringent starvation protein A"
/protein_id="AA066944.1"
/db_xref="GI:24982777"
KLAENPYGVPVLVDRLALYESTVVMYLEERYPHPLMPVYPVARGNSRLMHRI

```

QRDWCALADTVLDRSSEARTEARKALRESLTGVSPLFSEFACFMSDEQSLVDCCLL  
 PITMLRPLVLGTLPQAKPLLDYMERQAFREPFQASLSVEREMKL"  
 693..1154  
 /gene="sspb"  
 /locus\_tag="ppl321"  
 693..1154  
 /gene="sspb"  
 /locus\_tag="ppl321"  
 /note="similar to GB:S50157, GB:X60221, SP:P24539, and  
 PID:509291; identified by sequence similarity; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="stringent starvation protein B"  
 /protein\_id="AAN66945.1"  
 /db\_xref="GI:24982778"  
 /translation="MKCAFSEKPLMSSRPYLVRALYEMIVDNDCTPHMLYNAEPKV  
 QVPDGFASDGOIVLINISPAVSRLHMDNDVSEFGRFSGVAHSLFPVPGAILGIYARE  
 NGQGVFLEFSSMDGELEDDVQPDGDDGPDGQPPRTGRPSLWVK"  
 complement(1216..11794)  
 /locus\_tag="ppl322"  
 complement(1216..1794)  
 /locus\_tag="ppl322"  
 /codon\_start=1  
 /transl\_table=11  
 /product="lipoprotein, putative"  
 /protein\_id="AAN66946.1"  
 /db\_xref="GI:24982779"  
 /translation="MTPMRGLMALTLCLSVTGCSSVLTSTNSPIEDDRGTRTIGSK  
 IDDSLIEKAAVNTSKANPDKSHIVVSSYNGIVLLAGOTPRADLSLAERQASOV  
 QRKVVHQLQVMPSSILARNNDNLTKTKTOMLTDSAVFSSRIKVTENGIVILL  
 GLVTRQANSATAVQGVSGQIVKLFEYID"  
 complement(1791..2378)  
 /gene="gmhA"  
 /locus\_tag="ppl323"  
 complement(1791..2378)  
 /gene="gmhA"  
 /locus\_tag="ppl323"  
 /note="identified by match to PFAM protein family HMM  
 PF01380"  
 /codon\_start=1  
 /transl\_table=11  
 /product="phosphoheptose isomerase"  
 /protein\_id="AAN66947.1"  
 /db\_xref="GI:24982780"  
 /translation="MQSRIRLFERASITKQAMDILAPHIEQASLVMVNALLNEGKM  
 LACNGSGAGAHFSELLRERERPSLAIALTTDSITSTANDSYNEVFSKO  
 IRLAQGDVLLATISGNSAVIQAIQAHDREMIYVALTRDGGGNASLLLPEDVE  
 IRLPSTVTRIQEVHLLAIHCLDLIDSQLFSEE"  
 complement(2471..2845)  
 /locus\_tag="ppl324"  
 complement(2471..2845)  
 /locus\_tag="ppl324"  
 /note="identified by match to TIGR protein family HMM  
 TIGR00252"  
 /codon\_start=1  
 /transl\_table=11  
 /product="conserved hypothetical protein TIGR00252"  
 /protein\_id="AAN66948.1"  
 /db\_xref="GI:24982782"  
 /translation="FMAASPTRAGQAAEQALEYLOGQLQLLARNWRCKGGELDLV  
 MLDATVYFVVRVRLHAGFGGALDSIDGRKQKRLVLAATLFLQEAHWGNYPCRFDV  
 VALQGSFHAGRPLOWLNKAFEC"  
 complement(2842..4659)  
 /locus\_tag="ppl325"  
 complement(2842..4659)  
 /locus\_tag="ppl325"  
 /codon\_start=1  
 /transl\_table=11  
 /product="lipoprotein, putative"  
 /protein\_id="AAN66949.1"  
 /db\_xref="GI:24982783"  
 /translation="MIACLRLTLALCLAAALAACASSPSSSISGELPRTPPDASISQELLE  
 KAASSKAEDALLURLSAADLAYKQKQFPFRAARILEQVPLDTLPAQOVFASTLADEL

AMSRNQPKAALTALAHPSLQRLAEPLADLQARTYSVHAAALPADQALAAAOQRILLA  
 PLLGQAASNTNDIAIWLAVSPAELQLOPTNDQTLAGTSLAFVKSAGTLEQQQAA  
 IEAVVKQHPDHPAAQOLPLATLTKKELASOPTLTKLALLPOEGPLAGVARLUDGFMA  
 RHFOAQSGQAPAVQVPSRSSIGSLDDFYBQAQAAGVQLVIGPLEKPLVKOLAAPQ  
 LPTITLALNDAGQAKVPPQFQGLAAEDAREVARRADGCMWRANVALVSGEWGD  
 RYLAARQNDWEGNGTLLAERIAQFVALAQIADLFQLRQSEGRKASLQSVTVGSIA  
 AQPSRRQDIDFLASTPQQAQIKPTLNFQYAGDVPIYATSNLYSASGDIHQYNDMN  
 GTRFCETPWLDDTSNSLRQQVQWPAQAGSLGRLYAMGVDAVSLAPRLGQLKALPDN  
 RVGLSGSLSINANQVERQLPWAEEFSGQVKRLPDIVR"  
 4837..5721  
 /locus\_tag="ppl326"  
 4837..5721  
 /locus\_tag="ppl326"  
 /note="identified by match to PFAM protein family HMM  
 PF00590"  
 /codon\_start=1  
 /transl\_table=11  
 /product="tetracycline methylase family protein"  
 /protein\_id="AAN66950.1"  
 /db\_xref="GI:24982784"  
 /translation="MCAVTDVAGASKSTVGTLYVATPTGNLDDMSARALKVLADVAL  
 IAAEDTRHSIRLQHFQIDTPLAACHEHNERDEGGRFLSKLLAGENVALVSDAGTPLI  
 SDPGYHLVQRARAGSVVPPVPGACALIAALSAAGLPDSDFEFGFLPAKQARRARL  
 EQVREEPTLIFYRAPHRLILCEGDEMEAVFGDGRPALLARELTKTFFELKGLPLAELR  
 APVAGDSNQQRGECVVLVGGWSAPEGEQAISAEARVLDLLELLEPLKRAALAAEIT  
 GVRKNLLYQLALEKQAE"  
 complement(5744..6164)  
 /locus\_tag="ppl327"  
 /note="this region contains a gene with one or more  
 premature stops or frameshifts, and is not the result of a  
 sequencing artifact; conserved hypothetical protein,  
 degenerate; similar to GB:X16944, and SP:P15488;  
 identified by sequence similarity; putative"  
 6455..6928  
 /locus\_tag="ppl328"  
 6455..6928  
 /locus\_tag="ppl328"  
 /note="identified by match to PFAM protein family HMM  
 PF02381"  
 /codon\_start=1  
 /transl\_table=11  
 /product="conserved hypothetical protein TIGR00242"  
 /protein\_id="AAN66951.1"  
 /db\_xref="GI:24982786"  
 /translation="MWGITAVFRGANAVSLDAKGLAMPSTRYRDELDRCNCQLIVTI  
 DAVIDCLCVPLDDEWEIETAKLRALPSLEENRRLQLLIGNAVLELDGSGRFLVPP  
 RLREYAKDLKKMLVQLNKFQMDWEDANNAVSAADLAAAIQPGAMPDELRLIL"  
 6925..7872  
 /locus\_tag="ppl329"  
 6925..7872  
 /locus\_tag="ppl329"  
 /note="similar to GP:9950652; identified by sequence  
 similarity; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="conserved hypothetical protein TIGR00006"  
 /protein\_id="AAN66952.1"  
 /db\_xref="GI:24982787"  
 /translation="MTIDSGFNHITVLLDEAVEALRADGCYLDGTFGRGGHSRLIL  
 SKLPGQRLIGFDKDPQAIATGOALAAEDGRFIVQSRFAGELGAERGLGVKYSGL  
 LLDGVSSQDDDPERGSFNDGFLDMRNPDCQVSAEETATAPVEIARVFEYEG  
 EERFAGRNARAVERRERKQPTRTADLAELVKVANPEWKGKNPATRAPQGLRIHVN  
 ELGDLEAGLEAALDALEYGGRLLAVISFHLSDRIYVKLFMRILVKEADNLPRNLVQVH  
 KVFEPPKIKLIGKAQFASEALKANPRSRSAVMRVAEKL"  
 7869..8162  
 /gene="ftsL"  
 /locus\_tag="ppl330"  
 7869..8162  
 /gene="ftsL"  
 /locus\_tag="ppl330"  
 /note="similar to GB:M11186, GB:M62611, SP:P01178,  
 PID:L89411, PID:386991, and PID:825699; identified by  
 sequence similarity; putative"

## Alignment Scores:

Pred. No.: 5,36e-24 Length: 301995  
Score: 35.00 Matches: 35  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 7.81% Indels: 0  
DB: 1 Gaps: 0

US-09-701-229-2 (1-448) x AE016779 (1-301995)

QY 136 LysargValAlaValGlyAsnLeuGlyThrProAlaLeuAspLeuLeuAlaAspasp 155  
|||||  
Db 14248 AAACGCGTGGCGTGGGAGTAACTCGTACCGCGACCTGCTGCGCGATGAC 14307  
|||||  
QY 156 IleGluLeuTyrrValLeuGluLeuSerSerPheGlnLeuGluThr 170  
|||||  
Db 14308 ATCGAGCTGTAGTGTGGAGCTTTCAGCTCCAGCTCGGAACC 14352  
|||||

## RESULT 4

AE016871/c 311600 bp DNA linear BCT 05-MAR-2003  
LOCUS Pseudomonas syringae pv. tomato str. DC3000 section 16 of 21 of the  
DEFINITION complete genome.

ACCESSION AE016871 AE016853

VERSION AE016871.1 GI:28854552

KEYWORDS

SOURCE

ORGANISM

Pseudomonas syringae pv. tomato str. DC3000  
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
Pseudomonadaceae; Pseudomonas.

## REFERENCE

AUTHORS

1 (bases 1 to 311600)  
Buell, R., Joardar, V., Khouri, H., Fedorova, N., Tran, B., Russell, D.,  
Berry, K., Utterback, T., Van Aken, S., Feldblyum, T., Gwinn, M.,  
Dodson, R., DeBoy, R., Durkin, A., Kolonay, J., Madupu, R.,  
Daugherty, S., Brinkac, L., Beanan, M., Haft, D., Selengut, J.,  
Neilson, W., Davidson, T., White, O., Fraser, C. and Collmer, A.  
Complete Sequence of Pseudomonas syringae  
Unpublished

2 (bases 1 to 311600)

Buell, R., Joardar, V., Khouri, H., Fedorova, N., Tran, B., Russell, D.,  
Berry, K., Utterback, T., Van Aken, S., Feldblyum, T., Gwinn, M.,  
Dodson, R., DeBoy, R., Durkin, A., Kolonay, J., Madupu, R.,  
Daugherty, S., Brinkac, L., Beanan, M., Haft, D., Selengut, J.,  
Neilson, W., Davidson, T., White, O., Fraser, C. and Collmer, A.  
Direct Submission

TITLE

JOURNAL

Submitted (03-MAR-2003) The Institute for Genomic Research, 9712  
Medical Center Dr, Rockville, MD 20850, USA

## FEATURES

source

1..311600  
/organism="Pseudomonas syringae pv. tomato str. DC3000"  
/mol\_type="genomic DNA"  
/strain="DC3000"  
/db\_xref="taxon:223283"  
/note="pathovar: tomato"  
131..601  
/gene="bfr"  
/locus\_tag="PSPT04160"  
131..601  
/gene="bfr"

gene

131..601  
/locus\_tag="PSPT04160"  
/note="similar to SP:P22759, and SP:P22759; identified by  
sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="bacterioferritin"  
/protein\_id="AAO57616.1"  
/db\_xref="GI:28854553"

CDS

131..601  
/gene="bfr"  
/locus\_tag="PSPT04160"  
/note="similar to SP:P22759, and SP:P22759; identified by  
sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="bacterioferritin"  
/protein\_id="AAO57616.1"  
/db\_xref="GI:28854553"  
/translation="MKGDITVIQHLNKLIGNELVAINOYFLHARMYEDWGLKLGGEH  
YKESIHAMVADKLKRTFLFGLNPLNQLDGKILIGHTKEMLECDLKEOKGLADLK  
AAIAHFETVDFGSRLELLEKPEEQIDWLEVLQGLIVKVGNIYLOSQMGD"  
complement(689..1015)  
/locus\_tag="PSPT04161"

gene

CDS  
complement(689..1015)  
/locus\_tag="PSPT04161"  
/note="identified by match to TIGR protein family HMM  
TIGR00365"  
/codon\_start=1  
/transl\_table=11  
/product="glutaredoxin-related protein"  
/protein\_id="AAO57617.1"  
/db\_xref="GI:28854554"  
/translation="MDITETIKQIASNTILLYMKGAPNAPQCGPSAKASQALMACGE  
KFAYVDILQNFPEIRANLPKYANWTFPQLWAGELVGGSDIITENMADGSLQTLVKEA  
SAAKAE"  
complement(1104..3302)  
/locus\_tag="PSPT04162"  
complement(1104..3302)  
/locus\_tag="PSPT04162"  
/note="similar to GP:5441785; identified by sequence  
similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="oxidoreductase, molybdopterin-binding"  
/protein\_id="AAO57618.1"  
/db\_xref="GI:28854555"  
/translation="MARVIARHSHCSAIHRIAMTRPEAMTRTOHLRACHICEAICG  
LAIVTEPGAAPKITSIKGDPDPTFSRGHICPKAVALQDIONDRLRQMLRTGDQ  
WPIEMQAADFVAERLYAIQQHGNNAVYQGNPSVHNYGLMTHSNYFLGLLKTNR  
RFSATSDPLPHLTFLMYGHGMLLPIDIDHTDFMLILGNPLASNGSIMTVPDVE  
KRLKAIQORGKLVVDPRRSETAIAIOHVFVRPGDAALLGLLTLFEGLTRES  
HLPVDGLHVRHSIAGSFAEAMSPRCGLAIVQIQLARDPAGATACYGRMGVSTQA  
FGTICHLQAQLINLVNDRVGGTLCPEAVDLVSSGSGHFNWVRVSGLPEYGG  
ELPVSALAEMLGQGGQVRALVTVAGNPVLTSPNGRQLDALEGLFHLSTIDLYNE  
TTRHADLILPSTALENDHYDTFTNTLAVRNVTRFNRAIFDKPEGALHDETFVGLAA  
SFAKACORRELKTPVPAQMDIVLRAGRYGAASPHNLSLETLAGHPHGRNLSGKPNL  
TLRLKTAQRIOAAPEVTMADIVRAADGAPGCEKAGOLLIGRHRVRSNNWMMHY  
HRLYKGRPHQLLHPDDLHRGLSDGQQQVRVSRVGVIEVYELGSLDMQPGVSLPH  
GWGHSRSCVMEIARNQPGVSANDUTDERQLDELGSNAALNGVPVQVASC"  
complement(3357..3650)  
/locus\_tag="PSPT04163"  
complement(3357..3650)  
/locus\_tag="PSPT04163"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAO57619.1"  
/db\_xref="GI:28854556"  
/translation="MDSKEHOLTELLGLTARTLTLTASMTSMPELLRSEDEVTRNA  
GRMIDRMATISSGLDEHWRLLIGDLTGIHMAQEQVETVTEIQLQRKIYTPGG"  
4100..5020  
/gene="argF"  
/locus\_tag="PSPT04164"  
4100..5020  
/gene="argF"  
/locus\_tag="PSPT04164"  
/note="identified by match to PFAM protein family HMM  
PF02729"  
/codon\_start=1  
/transl\_table=11  
/product="ornithine carbamoyltransferase"  
/protein\_id="AAO57620.1"  
/db\_xref="GI:28854557"  
5026..6087  
/locus\_tag="PSPT04165"  
5026..6087  
/locus\_tag="PSPT04165"  
/note="similar to GB:248042, SP:Q14444, and PID:662994;  
identified by sequence similarity; putative"

```

/codon_start=1
/product="ABC transporter, ATP-binding protein"
/protein_id="AA057621.1"
/db_xref="GI:28854558"
/translation="MLNLRNLACGQYQNRVVDNLNHLNAGDTCGLLGGSCGKTTT
LRAAGFPIHAGETITLADGVSIRPGWTLAPKKRIGVMFODYALFPHLSVADNIAG
LRQPRLDVIGLELVLNLCALGKRVPHLSGGQQORVALARALAPPEQLLLDEFP
SNLQELRRRLSHVRDLKARGTSAILVTHDQBEAFVSDHGVGFVKGREQMDTPY
NLYHEPQTFVASIGQYFIRGQMIPEPSVTELVGLRNPYPYVGRGSAVDVLLRP
DDIVYAPSDLKAIRVIGRTFKGASTLYRLQLPTGSLEAIPFPPSHADHRQGEVGI
ADHLVFAAGSVAAYAG"
/complement(6036..6566)
/locus_tag="PSPT04166"
/complement(6036..6566)
/locus_tag="PSPT04166"
/notes="similar to SP:P21484; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="ybaK/ebcC protein"
/protein_id="AA057622.1"
/db_xref="GI:28854559"
/translation="MTPALDLLKKNRAEHRIHSYHDPKAAASYGLEAAEKLGDAARV
FKTLASTERKGLLVAVPVVGTLDLALANAAGAKKTENADPAAQRSTGYLLGGIS
PLGQKKRLRTFIDETALESIVSAGRGLELVETLTAVLAETHQARFAPIGRE"
6918..7775
/gene="glpP"
/locus_tag="PSPT04167"
6918..7775
/gene="glpP"
/locus_tag="PSPT04167"
/notes="similar to SP:P11244; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="glycerol uptake facilitator protein"
/protein_id="AA057623.1"
/db_xref="GI:28854560"
/translation="MTIALKQPTLCQCAEFLGTALMIFFGTGCVAALKVAGATFGL
WEICLWMAVSMGIVSAGISGAHLNPAVSIASLFAFGFKRKLFPYISAQIAGAF
GAGLVLYLISDFEHAHIIHQSESLASVFTSNPNAISVQAFLEVFWIT
ILMGVIMALDDNNGPLRGALAPLLGLLVAVIGSAMGPLTGFAMPARDPFGPKMTF
FAGWGIEATFTGGRDIPYFLVPIFAPILGACLAGAAYRIARLHPSTVTVENEKEAPV
VRGKVOASS"
7805..9310
/gene="glpK"
/locus_tag="PSPT04168"
7805..9310
/gene="glpK"
/locus_tag="PSPT04168"
/notes="identified by match to PFAM protein family HM
P02782"
/codon_start=1
/transl_table=11
/product="glycerol kinase"
/protein_id="AA057624.1"
/db_xref="GI:28854561"
/translation="MTDTQNKYIIIALDQGTSSRAIIFDRDANVYVSTAQSEFVQHPY
QAGVHDPMEIFATQCTACALQAADLHQAIDGITNQRTTVIWERDTGRPIY
NAIVQCRSTECQQLKRDLEEVYKDTTGLVLDLPYFSGSKVLLDHWGSSRERAR
KGLMFGITDWLWKFTGGKVHVVDYTNASPTMLENHTLEWDRMLDLDVLDPREIL
PVKSSSEVYGHKSQIPAGIAGQQAALFQCMQVEPQQAQNTGTGCLFLLMNTGKK
AVKSHGMLTTICGPRGEVATGALGVNGSTVQWLRDELKINDALDTEYFAGV
KDSNGYVLPATGLGAPWDPYARGALFGLTRGVKVDHIIIRAALESATYQTRDVLDA
MOQDSGERLKSURVDGAVANNFLMQFQADILGTHVERPQMRRTALGAFLAGLAIG
FWSSLDLNRKAVIERVEPACAEAREKLYAGWKQKAVARTDWEKPHENE"
9421..10176
/gene="glpP"
/locus_tag="PSPT04169"
9421..10176
/gene="glpP"
/locus_tag="PSPT04169"

```

/note="similar to SP:Q51391; identified by sequence similarity; putative"

Alignment Scores:  
 Pred. No.: 2.33e-08 Length: 311600  
 Score: 20.00 Matches: 20  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 4.46% Indels: 0  
 DB: 1 Gaps: 0

US-09-701-229-2 (1-448) x AE016871 (1-311600)

Qy 278 AsnAlaLeuAlaLeuAlaLeuGlyHisAlaValGlyLeuProPheAspAlaMetIeu 297  
 Db 291021 AACGACATGGCAGCCCTGGCGTCGGCATGCCCTGGCGCTTCCGACCCATGCTT 290962

#### RESULT 5

AX067460/c AX067460 96109 bp DNA linear PAT 24-JAN-2001  
 LOCUS AX067460 Sequence 35 from Patent WO0078968.  
 DEFINITION AX067460  
 ACCESSION AX067460  
 VERSION AX067460.1 GI:12545080  
 KEYWORDS Moraxella catarrhalis  
 SOURCE Moraxella catarrhalis  
 ORGANISM Moraxella catarrhalis  
 Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
 Moraxellaceae; Moraxella.

#### REFERENCE

1 Lagace,R.E., Patterson,C. and Berg,K.L.  
 Nucleotide sequences of moraxella catarrhalis genome  
 Patent: WO 0078968-A 35 28-DEC-2000;  
 Incyte Genomics, Inc. (US)

#### FEATURES

source  
 1..96109  
 /organism="Moraxella catarrhalis"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:480"

BASE COUNT 28783 a 18910 c 20341 g 28075 t  
 ORIGIN

#### Alignment Scores:

Pred. No.: 8.96e-08 Length: 96109  
 Score: 19.00 Matches: 19  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 4.24% Indels: 0  
 DB: 6 Gaps: 0

US-09-701-229-2 (1-448) x AX067460 (1-96109)

Qy 113 AlaIleThrGlySerAsnAlaLysSerThrValThrThrLeuValGlyGluMetAla 131  
 Db 12400 GCCATCACAGGCTCAATGCCAAGACACTGTTACGACATTAGTGGCGAGATGGCA 12344

#### RESULT 6

AX189058  
 LOCUS AX189058 1317 bp DNA linear PAT 08-AUG-2001  
 DEFINITION Sequence 259 from Patent WO0148209.  
 ACCESSION AX189058  
 VERSION AX189058.1 GI:15142764  
 KEYWORDS Escherichia coli  
 ORGANISM Escherichia coli  
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 Enterobacteriaceae; Escherichia.

#### REFERENCE

1 Forsyth,R.A., Ohlsen,K.L. and Zyskind,J.W.  
 Genes identified as required for proliferation of E. coli  
 Patent: WO 0148209-A 259 05-JUL-2001;  
 Elitra Pharmaceuticals, Inc. (US)

#### FEATURES

source  
 1..1317

CDS  
/organism="Escherichia coli"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:562"  
1. .1317  
/note="unnamed protein product"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAC50739.1"  
/db\_xref="GI:15142765"  
/translation="MADYQKNVILIGLTLGLSCVDFEFLARGVTPRVMDTRMTPPGL  
DKLEPEVERHTGSLNDEWMAADLIVASPGIALAHPSLSAADAGIEIVGDIELFCRE  
AQAPIVALTGSGKSTVTLVGENAKAAGVNVGGNIGLPLMLDDCELYVLELS  
SFQLETTSSLOAVATILNVTEDHMDRYPFGLQYRAAKLRIYENAKVCVYNADDAIT  
MPIGADRCVSVFVNMGDYHLNHOQGETWLRVGEKVLNVKEMKLSQHNVTNALAA  
LALADAGSLKALTFTGLPHRFVLEHNGVRIWINDSKATNVGSTEALNGL  
HVDGTLLHLLGGCKSADFSPLARYLNGDNVRLYCFGRDGAQALALRPEVAEQETME  
QAMRLAPRVQPGDMVLLSPACSLDQFKNEQRCNEFARLAKELG"

BASE COUNT 294 a 323 c 400 g 300 t

ORIGIN

Alignment Scores:  
Pred. No.: 0.000314 Length: 1317  
Score: 14.00 Matches: 14  
Percent Similarity: 100.00% Conservatives: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 3.12% Indels: 0  
DB: 6 Gaps: 0

US-09-701-229-2 (1-448) x AX189058 (1-1317)

QY 157 GluLeuTyRValLeuGluLeuSerPheGlnLeuGluThr 170

DB 457 GAACGTGACGTGCTGGAACCTCGAGCTTCAGCTGGAAC 498

RESULT 7

ECMURD 1629 bp DNA linear BCT 12-SEP-1993  
LOCUS Escherichia coli murd gene for  
DEFINITION UDP-N-acetylmuramoyl-L-alanine:D-glutamate ligase (EC 6.3.2.9).  
ACCESSION X17609  
VERSION X17609.1 GI:42045

KEYWORDS murd gene; peptidoglycan synthesis;  
UDP-N-acetylmuramoyl-L-alanine:D-glutamate ligase.

SOURCE Escherichia coli

ORGANISM Escherichia coli  
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
Enterobacteriaceae; Escherichia.

REFERENCE 1 (bases 1 to 442; 476 to 1629)

AUTHORS Mengin-Lecreux,D. and van Heijenoort,J.

TITLE Nucleotide sequence of the murd gene encoding the

JOURNAL UDP-MurNAc-L-Ala-D-Glu synthetase of Escherichia coli

MEDLINE Nucleic Acids Res. 18 (1), 183 (1990)

PUBMED 90174916

REFERENCE 2 (bases 1 to 1629)

AUTHORS Mengin-Lecreux,D.

TITLE Direct Submission

JOURNAL Submitted (23-NOV-1989) Mengin-Lecreux D., URA 1131 du CNRS,  
Biochimie Moleculaire et Cellulaire, Batiment 432, Universite Paris  
Sud, 91405 Orsay, France

COMMENT \*map: map position-2 minutes;

Data kindly reviewed (26-FEB-1990) by Mengin-Lecreux.

FEATURES

source

1. .1629

/organism="Escherichia coli"

/mol\_type="genomic DNA"

/strain="K12"

/db\_xref="taxon:562"

212. .216

/note="ribosome binding site"

227. .1543

/note="UDP-N-acetylmuramoyl-L-alanine:D-gluta ligase (AA 1-438)"

/codon\_start=1  
/transl\_table=11  
/protein\_id="CAA35611.1"  
/db\_xref="GI:42046"  
/db\_xref="SWISS-PROT:PI4900"  
/translation="MADYQKNVILIGLTLGLSCVDFEFLARGVTPRVMDTRMTPPGL  
DKLEPEVERHTGSLNDEWMAADLIVASPGIALAHPSLSAADAGIEIVGDIELFCRE  
AQAPIVALTGSGKSTVTLVGENAKAAGVNVGGNIGLPLMLDDCELYVLELS  
SFQLETTSSLOAVATILNVTEDHMDRYPFGLQYRAAKLRIYENAKVCVYNADDAIT  
MPIGADRCVSVFVNMGDYHLNHOQGETWLRVGEKVLNVKEMKLSQHNVTNALAR  
LALADAGSLKALTFTGLPHRFVLEHNGVRIWINDSKATNVGSTEALNGL  
HVDGTLLHLLGGCKSADFSPLARYLNGDNVRLYCFGRDGAQALALRPEVAEQETME  
QAMRLAPRVQPGDMVLLSPACSLDQFKNEQRCNEFARLAKELG"

BASE COUNT 351 a 403 c 486 g 389 t

ORIGIN

Alignment Scores:  
Pred. No.: 0.00038 Length: 1629  
Score: 14.00 Matches: 14  
Percent Similarity: 100.00% Conservatives: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 3.12% Indels: 0  
DB: 1 Gaps: 0

US-09-701-229-2 (1-448) x ECMURD (1-1629)

QY 157 GluLeuTyRValLeuGluLeuSerPheGlnLeuGluThr 170

DB 683 GAACGTGACGTGCTGGAACCTCGAGCTTCAGCTGGAAC 724

RESULT 8

ECMURD

LOCUS Escherichia coli murd gene and ORF-Y (EC 6.3.2.9).

DEFINITION X51584 X52117

ACCESSION X51584.1 GI:42058

VERSION UDP-N-acetylmuramoyl-L-alanyl-D-glutamate synthetase; unidentified

KEYWORDS reading frame.

SOURCE Escherichia coli

ORGANISM Escherichia coli

Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

Enterobacteriaceae; Escherichia.

REFERENCE 1 (bases 1 to 2608)

AUTHORS Ikeda,M., Wachi,M., Ishino,F. and Matsushashi,M.

TITLE Nucleotide sequence involving murd and an open reading frame ORF-Y

JOURNAL spacing murf and ftsW in Escherichia coli

MEDLINE Nucleic Acids Res. 18 (4), 1058 (1990)

PUBMED 90192099

REFERENCE 2 (bases 1 to 2608)

AUTHORS Ikeda,M.

TITLE Direct Submission

JOURNAL Submitted (02-JAN-1989) Ikeda M., Institute of Applied

Microbiology, The University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku,

Tokyo 113, Japan

REFERENCE 3 (bases 26 to 1354)

AUTHORS Flouret,B.

TITLE Direct Submission

JOURNAL Submitted (15-MAR-1990) Flouret B., URA 1131 du Centre National de

la Recherche Scientifique, C N R S, Biochimie Moleculaire et

Cellulaire, Batiment 432 Universite Paris-Sud, 91405 Orsay, France

COMMENT Sequence independently determined by [3], but authors accept

sequence [1] as definitive.

FEATURES Location/Qualifiers

source

1. .2608

/organism="Escherichia coli"

/mol\_type="genomic DNA"

/db\_xref="taxon:562"

42046



```
/note="ORF-Y (AA 1-360)"
/codon_start=1
/transl_table=11
/protein_id="CAA35932.1"
/db_xref="GI:42059"
/db_xref="SWISS-PROT:P15876"
/translation="MLYLAEHLVKYSGNVFSLTFRAIVSLTLTALFSLWNGPRM
IAHLQKLSQGVNRDGPESHFSKRGITPMGGIMLTAVISVLLWYPSNPYVVCVL
VWLVGYGVIFVDYDVRKDKTKGLIARKWYFMSVIALGVAFALYAGKDTPATOL
VWPFKDMQGLGYILLAYFVIVGTGNVNLTDGLDGLNIMPTVFVAGGFALVANA
TGNNFASYLHPIVLRHAGELVIVCTAIVGAGLGFLEWNTYPAQVFMGDVGSLSALGA
LGIITAVILROEFLVINGVVFETLSVILVQVSGFKLRGQIRFPMPIHHHYELKGHP
EPVIVFVWISLMVLIGLATLKV"
1018..1023
/note="pot. -35 region"
1043..1048
/note="pot. -10 region"
1181..1185
/note="pot. ribosome binding site"
1196..12512
/note="murD gene product (AA 1-438)"
/codon_start=1
/transl_table=11
/protein_id="CAA35933.1"
/db_xref="GI:42060"
/db_xref="SWISS-PROT:P14900"
/translation="MADYQKNVVIIGLTLGSCVDFFLARGVTPRVMDTRMTPPGL
DKLPEAVERTSLNDEWLMADLIIVASPGIALHPSLSAAADAGIEIVGDIELFCRE
AQAPIVAITSGNSKSTVTLVGMKAAGVNVGGNIGLPALMLDDECELVLELS
SFOLETTSSLOQAATILNVTEDHMDRYPFGLOQYRAAKLRIYENAKVCVNNADALT
MPTRGADERCVSGVANGDYHLNHQOGETWLRVKGKLVNKKEMKLSGQNVNNAALAA
LALADAGLPRASLKALITFTGLPHRFVLEHNGVRVMDKATNVGSTEALNGL
HVDTLHLGGGKGADFSPLARYLNDVRYCFGRGDAQLAALRPEVAEQETME
QAMRLAPRVQPDGVLVLLSPACSLQFKNFQGRNEFARLAKELG"
BASE COUNT 530 a 631 c 777 g 670 t
ORIGIN

Alignment Scores:
Pred. No.: 0.000579 Length: 2608
Score: 14.00 Matches: 14
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.12% Indels: 0
DB: 1. Gaps: 0

US-09-701-229-2 (1-448) x ECMUROY (1-2608)

Qy 157 GluLeuTyrValLeuGluLeuSerSerPheGlnLeuGluThr 170
|||||
Db 1652 GAACGTGACGTGCTGGAACGTGCGAGCTCCACCTCGAACC 1693

RESULT 9
AE015046
LOCUS
DEFINITION
Shigella flexneri 2a str. 301 linear BCT 18-OCT-2002
genome.
ACCESSION
AE015046 AE005674
VERSION
AE015046.1 GI:24050283
KEYWORDS
SOURCE
Shigella flexneri 2a str. 301
Shigella flexneri 2a str. 301
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Shigella.
1 (bases 1 to 11441)
REFERENCE
Jin,Q., Yuan,Z.H., Xu,J.G., Wang,Y., Shen,Y., Lu,W.C., Wang,J.H.,
Liu,H., Yang,J., Yang,F., Qu,D., Zhang,X.B., Zhang,J.Y., Yang,G.W.,
Wu,H.T., Dong,J., Sun,L.L., Xue,Y., Zhao,A.L., Gao,Y.S., Zhu,J.P.,
Kan,B., Chen,S.X., Yao,Z.J., He,B.K., Chen,R.S., Ma,D.L.,
Qiang,B.Q., Wen,Y.M., Hou,Y.D. and Yu,J.
Genome sequence of Shigella flexneri 2a: insights into
pathogenicity through comparison with genomes of Escherichia coli
K12 and O157
JOURNAL
Nucleic Acids Res. 30 (20), 4432-4441 (2002)
```

```
PUBMED 12384590
REFERENCE
2 (bases 1 to 11441)
AUTHORS
Jin,Q., Shen,Y., Wang,J.H., Liu,H., Yang,J., Yang,F., Zhang,X.B.,
Zhang,J.Y., Yang,G.W., Wu,H.T., Dong,J., Sun,L.L., Xue,Y.,
Zhao,A.L., Gao,Y.S., Zhu,J.P., Chen,S.X., Yao,Z.J., Wang,Y.,
Lu,W.C., Qiang,B.Q., Wen,Y.M. and Hou,Y.D.
Direct Submission
Submitted (08-MAY-2001) Microbial Genome Center of Chinese Ministry
of Public Health, 100 YingXin Jie, XuanWu Qu, Beijing 100052, P.R.
China
FEATURES
source
1..11441
/organism="Shigella flexneri 2a str. 301"
/mol_type="genomic DNA"
/strain="301"
/serotype="2a"
/db_xref="taxon:198214"
292..786
/genes="yabB"
/locus_tag="SF0078"
292..786
/genes="yabB"
/locus_tag="SF0078"
/notes="Residues 13 to 164 of 164 are 99 pct identical to
residues 1 to 152 of a 152 aa protein from Escherichia
coli O157:H7 EDL933 ref: NP_285777.1"
/codon_start=1
/transl_table=11
/product="orf, conserved hypothetical protein"
/protein_id="AA41743.1"
/db_xref="GI:24050284"
/translation="MMGVKVGIRGEAGMFGATLVNLDKGRLSVPTRYRQELLENAAG
QMLTIDIHQPCLLYLPPEWIEQLSRLLSMNPVRRVORLLGHASECOMDGAG
RLTIAPVLRHAGLTKEVLMVGQFNKELWDETTHQVQKEDIDAEQLATGDLSERLQ
DLSL"
788..1729
/genes="yabC"
/locus_tag="SF0079"
788..1729
/genes="yabC"
/locus_tag="SF0079"
/notes="Residues 1 to 313 of 313 are 99 pct identical to
residues 34 to 346 of a 346 aa protein from Escherichia
coli p1r: Q0ECF"
/codon_start=1
/transl_table=11
/product="yabC protein"
/protein_id="AA41744.1"
/db_xref="GI:24050285"
/translation="MMENYKHTTVLLDEAVNGLINIRPDGIYIDGTFGRGHSRLILSQ
LGEGRLLAIDRDPQAIYAKVTIDDPFRIIHGPFSAIGEYVAERDLKIGIDILDL
GVSSPOLDDAERGFSFMRDGPDMRMDPTRGSAAEWLQTAEEADIAWLKTYGEEHF
AKRIARAIVERNREQMTRTKELAEVVAATPVKDKFKHPATRTFOAIVRWNSELEE
IEQALKSSNLVAPGGLSLIISFHSLEDRIKREMSRSGQVPAGLPMTERQKLKL
GRQLRALKLMPEGEVAENPARSSVLRARIATNA"
1726..2091
/genes="ftsL"
/locus_tag="SF0080"
1726..2091
/genes="ftsL"
/locus_tag="SF0080"
/notes="Residues 1 to 121 of 121 are 100 pct identical to
residues 1 to 121 of a 121 aa protein from Escherichia
coli O157:H7 EDL933 ref: NP_285779.1"
/codon_start=1
/transl_table=11
/product="cell division protein; ingrowth of wall at
septum"
/protein_id="AA41745.1"
/db_xref="GI:24050286"
/translation="MISRVTEALSKVSGMSGHERHALPGVIGDDLRFGLKPLCLFI
CIIITAVTVTTAHTRLTLTAQREQVLRLERDALDIERNLILEENALGDHSRVERIAT
EXLQHQHVDPSQENIVVQK"
```





JOURNAL MEDLINE PUBMED REFERENCE AUTHORS	Science 269 (5223), 496-512 (1995) 95350630 7542800 2 (bases 1 to 11498) Tatusov,R.L., Mushegian,A.R., Bork,P., Brown,N.P., Hayes,W.S., Borodovsky,M., Rudd,K.E. and Koonin,E.V. Metabolism and evolution of Haemophilus influenzae deduced from a whole-genome comparison with Escherichia coli Curr. Biol. 6 (3), 279-291 (1996) 96398784 MEDLINE PUBMED REFERENCE AUTHORS TITLE JOURNAL	gene CDS	1585. .1908 /gene="H11131" 1585. .1908 /gene="H11131" /note="similar to SP:P22187 GB:K00137 GB:S49802 GB:S49875 GB:X52063 percent identity: 36.63; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="cell division protein (ftsL)" /protein_id="AAC22786.1" /db_xref="GI:1574686" /translation="MSENNKPRYPLOQILVDFSSNKLVLVLLIGLIVSAMGTIWT HKTRQLISENGLILQRALENEYRNQVQEAEGDSTRVESAISTAELMKVYSSSEQE VEIRE" 1921. .3753 /gene="H11132" 1921. .3753 /gene="H11132" /note="similar to GB:D10483 SP:P04286 GB:K00137 GB:S49802 GB:S49875 percent identity: 53.11; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="penicillin-binding protein 3 (ftsI)" /protein_id="AAC22787.1" /db_xref="GI:1574687" /translation="MVKFNSSRSKSGSKKTIKLTAPETVKONKPKVFEKCFMGRGY MLSTVILLGLCALVRAAYVOSINADTISNEADKRSKRDEVLSVRGSLDRNGQLL SVSPMSAIVADPKTKMLKNSLADKRIALAEELGEMTENDVKYKLEKSKNGQLL ROVELSKANYIRLKIKGIIETHEHFRFVVEAAHVGYTDIDNGTEGTEKFSNS LIVGKDGSTVRKGNIVAHISDEKDYDAQDVTLSIDEKLSQMYRIKKAIVSENN AEGGAVLVDRTEGLVAMATAPSVNPNRVGVKSELNRAITDFFERGTVKPFV LTAQORGVKRODEIITDTFSKISGKEIVDVAQAOCTLDEILMNSNRGVSLALRMP PSALMETYONAGLSKPTDLGLIGEQVILNANRKHADIERATVAIGYGITATPLQIA RAYALGSGYVRPISITKVPDPVIGKRVSEKIKDKVILEKVAIKNRKRAVVGIR VGVKTGTARKIENGHVNYKIVAFPTAGIAPISDPYALVVLINDPKRAGEYVGGAVSAPV FSNMGYALRANAIPQDAEAENETTTKSAKRIYVIGEHNKQVN" 3763. .5229 /gene="H11133" 3763. .5229 /gene="H11133" /note="similar to GB:D10483 SP:P22188 GB:X55814 PID:285768 PID:581032 percent identity: 55.72; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="UDP-N-acetylmuramyl-tripeptide synthetase (murE)" /protein_id="AAC22788.1" /db_xref="GI:1574688" /translation="MKKLTALFNLPELKNIDIELHNVLDSRKVKAGDLFVAIKGHOVD GNQFDSALHSGASAVSELTSESLTVAFTGNVYVVKYOLAHHLSSLDVDFVDS SNLTLVGTCTNGKTTISOLLAONAEHLGHPAAVMTIGNGLGQIVPAKNTGSAY ETQSLSAFKHAGADFTSIESSHGIAQHRVLAHFKAIFTNLDRHLDYHQSMENY AAAKRLTFELDTQIKVINADDEIQWLTPELDAIVASMNADFKVGSQHWKKAINH YHFGADITFESSGNGVLYHSLPGLAFNYSNLLVMTLLSFGYENLLATAKSLKG VCGRMETIYQPNKPTVYDVAHTPDALAKALIAAREHQGLWCIFGCGGDRGKR LMAQVQEAQFAEKIIVTKDNPTESQSQIETDIVAGFKNKEKVGIIIPDRAQAQIAIES AVENDVILAKGHEHYQIIGSEVVFHFSQDEALDFLK" 5243. .6616 /gene="H11134" 5243. .6616 /gene="H11134" /note="similar to GB:D10483 SP:P11880 PID:216500 PID:40854 PID:42048 percent identity: 51.88; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="UDP-MurNac-pentapeptide synthetase (murF)" /protein_id="AAC22789.1" /db_xref="GI:1574689" /translation="MKLSTVQLAQILQAKLIGDENVOQKINTDTRKSVSNLSLFFAL
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE JOURNAL REMARK	White,O., Clayton,R.A., Kerlavage,A.R. and Fleischmann,R.D. Direct Submission Submitted (25-JUL-1995) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA 4 (bases 1 to 11498) White,O., Clayton,R.A., Kerlavage,A.R. and Fleischmann,R.D. Direct Submission Submitted (27-SEP-1997) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA The H. influenzae sequence has been updated by R. Fleischmann. New database matches have been assigned, product names have been improved, and a number of frame shifts have been corrected. We gratefully acknowledge the work of Tatusov et. al. We have incorporated their annotation into the /notes fields of the corresponding H. influenzae genes 5 (bases 1 to 11498) White,O., Clayton,R.A., Kerlavage,A.R., Fleischmann,R.D., Peterson,J., Hickey,E., Dodson,R. and Gwinn,M. Direct Submission Submitted (28-MAY-1998) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA The whole genome was shifted by 588 nucleotides for a new start On Oct 1, 1996 this sequence version replaced gi:1221887. Location/Qualifiers 1. .11498 /organism="Haemophilus influenzae Rd" /mol_type="genomic DNA" /db_xref="taxon:71421" 131. .586 /gene="H11129" 131. .586 /gene="H11129" /note="similar to SP:P22186 GB:X52063 PID:40849 PID:42319 GB:U00096 percent identity: 59.21; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="AAC22784.1" /db_xref="GI:1574684" /translation="MFRGATVNLDSKGRVAIPTRYRAILEKNOQMVCVTVDIRQSC LLIYPLDEWEKTEQLLALSDFPTQRRQRLQVLMGHATECENDAOGRILLSGPLQHA KLEKGLMLVQLNKFISWDSVWHQIAEDIEIGSSTDFAADALNDFSL" 617. .1582 /gene="H11130" 617. .1582 /gene="H11130" /note="similar to GB:D10483 SP:P18595 GB:X52063 PID:216497 PID:40850 percent identity: 62.09; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="AAC22785.1" /db_xref="GI:1574685" /translation="WNSSENSFSSSEHITVLLHEAVNGLKENGIIIDGTGRCGHRS FILSQLSSNRLIADVDRPRAEAHKIQDLRFQHSFHSHPICDLNVLGVKIDG IILDLGSSPQDEARGFSGFMKDGFLDMRDTQGLSAEELKQVSTEDLTWLKTF GERFAKRIATAIVDFNKSVAVNGTEFLSRTSQLAELISQAVPFKDKHKHPATRSFOA TRIFNSELDELSLNSALOMLAPEGRLSISFSLSDRMVHKPMKQSGEDIPKG LPLREDQIQRNQKRLIICKAIQPSDAEIQANPRSRSAILRAERI"	gene CDS	gene CDS

KGEKFDHQYLDQVSGALVAVVQOENSSISVPQLVVKDTRIALGELAKWLREKINP  
RTVAMTSSGTTTKENTASILTQHTAADSEAVLFTNGFNNDIGVPLTLRLTEKHRF  
AIVELGANHONEYITTKLQPNAAIINNAPAHLEGFAGVVOAKGTYIRGLTKN  
GVATINAEHLDDIWOKEISNHAIOYENGKDISAKNIHTTSGSTFTLISPOGEIET  
LPYLGEHVNKALATAMVGNLTLDVAKGLEQRSQVKGRLFPPIQVTPNLLIDDT  
YNANKDSICRAIDVLKGYDAPRILCVGDMKELGENSLAIHREVGQYINLVNDDLCVSY  
GNESAVTSEAVSGKHFTDKTEMVDPLVPLIENQLQNNKVVVLGKGRSRMKMEDVIYS  
LKDKIKC

gene

6610.7692

/gene="H1135"

CDS

6610.7692

/gene="H1135"

/note="similar to GB:D10483 SP:PI5876 GB:X51584 PID:216501  
PID:40855 percent identity: 76.67; identified by sequence  
similarity: putative"

/codon\_start=1

/transl\_table=11

/product="phospho-N-acetylmuramoyl-pentapeptide-

transferase E (mray)"

/protein\_id="AAC22790.1"

/db\_xref="GI:1574690"

/translation="MLVLAELVRYETAFAISYITVRANLALLTALFISLWIGPKV  
IKRLQILKFGQEVNDPESHFAKKTPTMGVYMLFSIGVSTLLNANLPIYIWLCL  
FVLFGAIGFVDDFRITKRNITDGLIARNVFNWVVALVAILMLVWGLDHTDTRL  
VIFPKDIMPQLGLFYIVLSYFVIVGTGNVNLTDGLDGLAIMPALVWAGAFALIAWA  
TGNVFAEYLIHPYIKYSSVVEFTAIVGASGLFWNTYPAQVFMGVDVGSALGGA  
LGVVAILVROEFLVIMGVFVVEALSIVLQVGSYKLKQKQIFRMADTHHFFELKGWP  
EPRVIRFIWISLMLVLMGLVTLKL"

gene

7815.9128

/gene="H1136"

CDS

7815.9128

/gene="H1136"

Alignment Scores:

Pred. No.: 0.00218 Length: 11498  
Score: 14.00 Matches: 14  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 3.12% Indels: 0  
DB: 1 Gaps: 0

US-09-701-229-2 (1-448) x U32793 (1-11498)

Qy 157 GluLeuTyrValLeuGluSerSerPheGlnLeuGluThr 170

Db 8271 GAACCTTATGCTAGAGCTTCTAGTTTTCAGCTTGAGACA 8312

RESULT 11

AX191762

LOCUS

AX191762 Sequence 44 from Patent WO0149775. linear PAT 15-AUG-2001

DEFINITION

AX191762

ACCESSION

AX191762.1 GI:15209931

VERSION

KEYWORDS

Haemophilus influenzae

Haemophilus influenzae

ORGANISM

Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;

Pasteurellaceae; Haemophilus.

REFERENCE

Iversen, P.L.

Antisense antibacterial cell division composition and method

JOURNAL

Patent: WO 0149775-A 44 12-JUL-2001;

Avi Biopharma, Inc. (US)

FEATURES

Location/Qualifiers

1..11498

/organism="Haemophilus influenzae"

/mol\_type="genomic DNA"

/db\_xref="taxon:727"

BASE COUNT 3446 a 1842 c 2509 g 3701 t

ORIGIN

Alignment Scores:

Pred. No.: 0.00218 Length: 11498

Score: 14.00 Matches: 14  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 3.12% Indels: 0  
DB: 6 Gaps: 0

US-09-701-229-2 (1-448) x AX191762 (1-11498)

Qy 157 GluLeuTyrValLeuGluSerSerPheGlnLeuGluThr 170

Db 8271 GAACCTTATGCTAGAGCTTCTAGTTTTCAGCTTGAGACA 8312

RESULT 12

AE005185

LOCUS

DEFINITION

Escherichia coli O157:H7 EDL933 genome, contig 1 of 3, section 9 of

155.

ACCSSION

AE005185 AE005174

VERSION

AE005185.1 GI:12512782

KEYWORDS

SOURCE

ORGANISM

Escherichia coli O157:H7 EDL933

REFERENCE

AUTHORS

1 (bases 1 to 12791)

Perna, N.T., Plunkett, G. III, Burland, V., Mau, B., Glasner, J.D.,

Rose, D.J., Mayhew, G.F., Evans, P.S., Gregor, J., Kirkpatrick, H.A.,

Posfai, G., Hackett, J., Klink, S., Boutin, A., Shao, Y., Miller, L.,

Grotbeck, E.J., Davis, N.W., Llm, A., Dimalanta, E., Potamouis, K.,

Apodaca, J., Anantharaman, T.S., Lin, J., Yen, G., Schwartz, D.C.,

Welch, R.A. and Blattner, F.R.

TITLE

JOURNAL

MEDLINE

PUBMED

2 (bases 1 to 12791)

Perna, N.T., Plunkett, G. III, Burland, V., Mau, B., Glasner, J.D.,

Rose, D.J., Mayhew, G.F., Evans, P.S., Gregor, J., Kirkpatrick, H.A.,

Posfai, G., Hackett, J., Klink, S., Boutin, A., Shao, Y., Miller, L.,

Grotbeck, E.J., Davis, N.W., Llm, A., Dimalanta, E., Potamouis, K.,

Apodaca, J., Anantharaman, T.S., Lin, J., Yen, G., Schwartz, D.C.,

Welch, R.A. and Blattner, F.R.

Direct Submission

Submitted (22-OCT-2000) Laboratory of Genetics, University of

Wisconsin, 445 Henry Mall, Madison, WI 53706, USA

FEATURES

Location/Qualifiers

1..12791

/organism="Escherichia coli O157:H7 EDL933"

/mol\_type="genomic DNA"

/strain="EDL933"

/serotype="O157:H7"

/db\_xref="taxon:155864"

/note="enterohemorrhagic"

72..1076

/gene="fruR"

/note="synonym: Z0090"

72..1076

/gene="fruR"

/function="regulator; Energy metabolism, carbon:

Glycolysis"

/note="Residues 1 to 334 of 334 are 100.00 pct identical

to residues 1 to 334 of 334 from Escherichia coli K-12

Strain MG1655: B0080"

/codon\_start=1

/transl\_table=11

/product="transcriptional repressor of fru operon and

others"

/protein\_id="AAG54384.1"

/db\_xref="GI:12512783"

/translation="MKLDEIARLAGVSRRTASYINGKAKOYRVSDKTVKVMVYRE

HNYPHNAAGIRAGRTSGLIVPDLENTSYTRIANYLERQARGVQLIACSEDO

PDNEMRCIEHLQVQDAILIIVSTSLPPEHPTFORWANDPFFIVALDRDREHFTSVV

GADQDDAEMLAEURKFPATYVLYLGALFELSVSFLREQGFTAKWKDDPPEVHFYLYAN

```

SYREAAQLEFEKLETHMPQALFTTTSFALQGVMDVTLRRDGLKPLSDLAIAATFGDN
ELLDFLQCPVLVAQRHRDVAERVLEIVLASLDEPRKPKGLTRIKRNLRRGVLSRS
1678. .2136
/gene="yabB"
/note="synonym: Z0091"
CDS
1678. .2136
/gene="yabB"
/function="orf; Unknown function"
/note="Residues 1 to 152 of 152 are 99.34 pct identical to
residues 1 to 152 of 152 from Escherichia coli K-12 Strain
MG1655; B0081"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAG54385.1"
/db_xref="GI:12512784"
/translation="MFGATLVNLDKGRSLVPTRYREQLLENAAQOMVCTIDIIHPC
LLYPLPELIEIKSLRSLSSMPVRRVORLLGHASEQMDGAGRLIAPVLRQHA
GLTKEVNLGQFNKFLMDTETWHQOVKEDIDAEQLATGDLSERLQDLSI"
2138. .3079
/gene="yabC"
/note="synonym: Z0092"
CDS
2138. .3079
/gene="yabC"
/function="orf; Not classified"
/note="Residues 1 to 313 of 313 are 100.00 pct identical
to residues 1 to 313 of 313 from Escherichia coli K-12
Strain MG1655; B0082"
/codon_start=1
/transl_table=11
/product="putative apolipoprotein"
/protein_id="AAG54386.1"
/db_xref="GI:12512785"
/translation="WMENKHTTVLLDEAVNGLIRPDGIYIDGTFGRGHSRLILSQ
LGEGRDLAIDRPOAIAVAKTIDDPRESIINGPFSALGEYVAERDLKIDGILLDL
GVSSQLDDAERGFSPMRDGLDMDFRPGSAAEWLQTAEEADIAWLVKTYGGEERF
AKRIARAVERNEQPMRTKELAEVAAATPVKDKFHPATRTFQAVRIWYNSLEE
ITQALKSSNLAPGRLSIISFHSLEDRIVRKFRNRSRGQVPQVAGLPMTEQLKLL
GQRQLRGLKMPGEEVAENPRARSSVLRTAERTNA"
3076. .3441
/gene="ftsL"
/note="synonym: Z0093"
CDS
3076. .3441
/gene="ftsL"
/function="phenotype; Cell division"
/note="Residues 1 to 121 of 121 are 100.00 pct identical
to residues 1 to 121 of 121 from Escherichia coli K-12
Strain MG1655; B0083"
/codon_start=1
/transl_table=11
/product="cell division protein; ingrowth of wall at
septum"
/protein_id="AAG54387.1"
/db_xref="GI:12512786"
/translation="MISVTLSKVKGSMGSHERHALPGVIGDDLRLRFGKLPICLFI
CIIITAVTVTTHHTRLTLTAQREQLVLERDALDIEWRNLLEENALGDHRSVERIAT
EKLQKHQVDPQSQENIVQK"
3457. .5223
/gene="ftsI"
/note="synonym: Z0094"
CDS
3457. .5223
/gene="ftsI"
/function="enzyme; Cell division"
/note="Residues 1 to 588 of 588 are 100.00 pct identical
to residues 1 to 588 of 588 from Escherichia coli K-12
Strain MG1655; B0084"
/codon_start=1
/transl_table=11
/product="septum formation; penicillin-binding protein 3;
peptidoglycan synthetase"
/protein_id="AAG54388.1"
/db_xref="GI:12512787"

```

Alignment Scores:  
Pred. No.: 0.00239  
Score: 14.00  
Percent Similarity: 100.00%

```

/translacion="MKAANKQPKRQEEHANSISWRPALLCGCILLALAFILGRVAM
LQIVSDMLKREGDMSLRVQOVSTSRGMITDSGRPLAVSPVPKAIWADPKVEYHDAG
GISVGDWKALANALNIPDLQASINARINAKGRFIYLARQVNPDMADYIKKLKLPGLH
LREESRRYPSGVAHTLIGFTNVDSQIEGVEKSFQKLTGQGERIVKDRYGRVI
EDISSYDSQAHHNLSIDRLQALVYRELNNAFAFNKAEGSAVLVDVNTGEVLAMA
NSPYNNLSGTPKEMRNRTITDVEPGSTVKPVMYATALQGVVRENSVLTPIY
RINGHEIKVARYSELTLTGVLQKSSNVGSKLAMPSSALVDYTSRFGLGKATNIG
LVGERSGLYPQKORWSDIERATESFGGLMVTPLQLARVATIGSYIYRPLSITKVD
PPVGERVPESIVRTVHMMSVALPGGGVKAARKYIAIKTKTAKKVGPDGRVI
NKYIATAGYAPASQRFALVYVINDPQAGKYGGAVSADVFAGIMGVLTURTNIEPD
ALTGDKNEFVINGQEGTGGRS"
5210. .6697
/gene="murE"
/note="synonym: Z0095"
CDS
5210. .6697
/gene="murE"
/function="enzyme; Cell envelop: Murein sacculus,
peptidoglycan"
/note="Residues 1 to 495 of 495 are 98.98 pct identical to
residues 1 to 495 of 495 from Escherichia coli K-12 Strain
MG1655; B0085"
/codon_start=1
/transl_table=11
/product="meso-diaminopimelate-adding enzyme"
/protein_id="AAG54389.1"
/db_xref="GI:12512788"
/translation="MADRNLRLDLPWPDPAPSRALREMTLDSRVAAGDLFVAVVGH
QADRRYIPQAIAGVAAIAEAGEATDGEIREMHGVYIYLSQNLERLSLAGRFY
HPSDNLXXVGTGTNGTITQQLAQMSQLLGETSAVMGTGVLGKVIPTENTTG
SAVDQHELGLVDQDATCAMESVSHGLVQHRVAALKFAASVETNLSRDHLDYHCDM
EHYEAKWLLYSEHHCGQAIINADDEVRMLKLPDPAVAVSMEHINPCHGRWLKA
TEVNYHDSGATIRFSSWGDGIESHLGAFNWSNLLALALGLYPLADLLTAA
RLQPCGRNEVETAFGKPTVVVDYATPDPALEKALQAAHLCAHGLWCYFGCGDQDK
GRPLATAEEFADVAVVTDNPRTEPEPRAINDILAGMLDAGHAKVMEDRAEAVTC
AVMQAKENDVVLVAGKHEDYQIVGNQRLDYSRVTVVARLLGVIA"
6694. .8052
/gene="murF"
/note="synonym: Z0096"
CDS
6694. .8052
/gene="murF"
/function="enzyme; Cell envelop: Murein sacculus,
peptidoglycan"
/note="Residues 1 to 452 of 452 are 99.33 pct identical to
residues 1 to 452 of 452 from Escherichia coli K-12 Strain
MG1655; B0086"
/codon_start=1
/transl_table=11
/product="D-alanine-D-alanine-adding enzyme"
/protein_id="AAG54390.1"
/db_xref="GI:12512789"
/translation="MISVTLSQLTDLNGELQAGDITLDVATTTRKLTGCLFVALK
GERDHDQAKAGGAGALLVSRPLDIDLPQIVKDKTRLAFGELAAWVQQVPARY
VALTGGSSGSKYKEMTAAILSQCNLTLYTAGNLNNDIGVPMTLRLTPEYDYAVIELG
AHQGEIVTVTLTRPEALVNNLAAHLEGFGLAGVAKAKGFEISGLPENGIAIMN
ADNDWLNQSVIGSRKVRWFRSPNAANSDFATNTHVTSCHTEETLQTPGTSVDVLLP
LPGRHNTANALAAALSNSVCATLDALKAGLANLKVPGRLFFIQLAENQLLDDSYL
ANVGSMTAAQVQLAEMFGYRVLVVGDMALGESEACHVQVGEAAKQDRLVLSVGK
QSHAISTAGVGVEHFAEKTILTRKLSLIAEQVITILVKGSRSAAMEVVVALQENG
TC"
8046. .9128
/gene="mraY"
/note="synonym: Z0097"
CDS
8046. .9128
/gene="mraY"
/function="enzyme; Cell envelop: Murein sacculus,
peptidoglycan"
/note="Residues 1 to 360 of 360 are 99.72 pct identical to
residues 1 to 360 of 360 from Escherichia coli K-12 Strain
MG1655; B0087"

```

Alignment Scores:  
Pred. No.: 0.00239  
Score: 14.00  
Percent Similarity: 100.00%

Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 3.12% Indels: 0  
 DB: 1 Gaps: 0

US-09-701-229-2 (1-448) x AE005185 (1-12791)

Qy 157 GluLeuTyrValLeuGluLeuSerPheGlnLeuGluThr 170  
 |||||  
 Db 9587 GAAGTGTACGTACTGGAAGTCTGAGCTTCAGCTGGAAACC 9628

## RESULT 13

AE000118

LOCUS

DEFINITION Escherichia coli K12 MG1655 section 8 of 400 of the complete genome. BCT 01-DEC-2000

ACCESSION AE000118 U00096

VERSION AE000118.1 GI:1786262

KEYWORDS

SOURCE ORGANISM

Escherichia coli K12  
 Escherichia coli K12  
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 Enterobacteriaceae; Escherichia.

## REFERENCE

AUTHORS

1 (bases 1 to 21757)  
 Blattner, F.R., Plunkett, G. III, Bloch, C.A., Perna, N.T., Burland, V.,  
 Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F.,  
 Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J.,  
 Mau, B. and Shao, Y.

The complete genome sequence of Escherichia coli K-12

Science 277 (5331), 1453-1474 (1997)

7426617

9278903

## REFERENCE

AUTHORS

Blattner, F.R.

Direct Submission

Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,  
 University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.

Email: ecollegenetics.wisc.edu Phone: 608-262-2534 Fax:

608-263-7459

3 (bases 1 to 21757)

Blattner, F.R.

Direct Submission

Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,  
 University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.

Email: ecollegenetics.wisc.edu Phone: 608-262-2534 Fax:

608-263-7459

4 (bases 1 to 21757)

Plunkett, G. III.

Direct Submission

Submitted (13-OCT-1998) Laboratory of Genetics, University of

Wisconsin, 445 Henry Mall, Madison, WI 53706, USA

This sequence was determined by the E. coli Genome Project at the  
 University of Wisconsin-Madison (Frederick R. Blattner, director).  
 Supported by NIH grants HG00301 and HG01428 (from the Human Genome

Project and NCHGR). The entire sequence was independently  
 determined from E. coli K12 strain MG1655. Predicted open reading

frames were determined using GeneMark software, kindly supplied by  
 Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA,

30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that

have been correlated with genetic loci are being annotated with CG

Site Nos., unique ID nos. for the genes in the E. coli Genetic

Stock Center (CGSC) database at Yale University, kindly supplied by

Mary Berlyn. A public version of the database is accessible

(http://cgsc.biology.yale.edu). Annotation of the genome is an

ongoing task whose goal is to make the genome sequence more useful

by correlating it with other data. Comments to the authors are

appreciated. Updated information will be available at the E. coli

Genome Project's World Wide Web site

(http://www.genetics.wisc.edu). \*\*\* The E. coli K12 sequence and  
 its annotations are periodically updated; this is version M54. No  
 sequence changes. Annotation updates: updated gene identifications

and products; all new functional assignments courtesy of Monica  
 Riley; added promoters, protein binding sites, and repeated  
 sequences described in reference 1. The unique numeric identifiers

## FEATURES

source

1. .21757

/organism="Escherichia coli K12"

/mol\_type="genomic DNA"

/strain="K12"

/sub\_strain="MG1655"

/db\_xref="taxon:83333"

complement(77..163)

/gene="leuL"

/note="synonym: b0075"

complement(77..163)

/gene="leuL"

/function="leader; Amino acid biosynthesis: Leucine"

/note="f28; 100 pct identical to LPL\_ECOLI SW: P09149"

/codon\_start=1

/transl\_table=11

/product="leu operon leader peptide"

/protein\_id="AAC73186.1"

/db\_xref="GI:1786263"

/translation="MTHIVRFIGLLLLNASSLRGRVSGIQH"

complement(197..226)

/note="factor Sigma70; promoter leu; documented +1

at83735"

complement(201..221)

/note="No predicted promoter"

/bound\_moiety="LexA predicted site"

complement(201..221)

/note="central position to predicted promoter: -0.5"

/bound\_moiety="LexA predicted site"

complement(354..368)

/note="No predicted promoter"

/bound\_moiety="Lrp documented site"

complement(354..368)

/note="central position to leuABCD promoter: -156.5"

/bound\_moiety="Lrp documented site"

460..474

/note="central position to predicted promoter: -28.5"

/bound\_moiety="Lrp predicted site"

462..490

promoter

/note="factor Sigma70; predicted +1 start at 84042"

complement(519..537)

/note="central position to predicted promoter: -319.5"

/bound\_moiety="TyrR predicted site"

554..565

/note="central position to predicted promoter:64"

/bound\_moiety="AraC predicted site"

646..1767

gene

/gene="leuO"

/note="synonym: b0076"

646..1767

/gene="leuO"

/function="putative regulator; Amino acid biosynthesis:

Leucine"

/note="o373; 100 pct identical to LEUO\_ECOLI SW:

P10151(290 aa) but contains 59 additional N-term aa and 24

C-term residues; 100 pct identical to PIR: S40589 but

contains 24 additional C-term residues"

/codon\_start=1

/transl\_table=11

/product="probable transcriptional activator for leuABCD

operon"

/protein\_id="AAC73187.1"

/db\_xref="GI:1786264"

/translation="MTHSTAMDSVFTRIFMFSEYFCFFLFYMHDKSYSSGLFIC  
 IPIRELSVTVLSMPEVOTDHPETAELSKPOLRMVDLNLTFVDAMQONITRAA  
 HVLGMSQPAVSNARLKVNFDELFRYGRGIQPTARAFQLFSGVFOALQVONELP  
 GSGFEPASSERVFHLCVCSPLDSILISQIYNHIEQIAPNHYVFKKSLNONTHERLY  
 QETFEVISEDHFRPEFTSVPLFKDEMVLVASKNHTPIKGLPLKHVDVYNQHAASVLD  
 RFASFSPQWYTDVKQASTAYQGMAMSVLSVSQTHLVAIAPRWLAEEFAESLELQV

beginning with a lowercase 'b' assigned to each gene (protein- or  
 RNA-encoding) are now designated as gene synonyms instead of  
 labels. This should allow them to be searched for in Entrez as gene  
 names.

```

protein_bind
1802..1816
/Note="central position to predicted promoter: -39.5"
/bound_moiety="Lrp predicted site"
1813..1839
/Note="factor Sigma70; promoter ilvIHp4; documented+1 at
85394"
1833..1847
/Note="central position to predicted promoter: -8.5"
/bound_moiety="Lrp predicted site"
1838..1866
/Note="factor Sigma70; promoter ilvIHp3; documented+1 at
85430"
1915..1929
/Note="central position to predicted promoter:73.5"
/bound_moiety="Lrp predicted site"
1949..1963
/Note="central position to predicted promoter:107.5"
/bound_moiety="Lrp predicted site"
1955..1984
/Note="factor Sigma70; promoter ilvIHp2; documented+1 at
85534"
1978..1992
/Note="central position to predicted promoter:136.5"
/bound_moiety="Lrp predicted site"
1995..3809
/Note="synonym: b0077"
/EC_number="4.1.3.18"
/EC_number="4.1.3.18"
/function="enzyme; Amino acid biosynthesis: Isoleucine,
Valine"
/Note="o604; 98 pct identical (1 gap) to 522 residues
from ILVI ECOLI SW: P00893 (566 aa)"
/codon_start=1
/transl_table=11
/product="acetylcofate synthase III, valine sensitive,
large subunit"
/protein_id="AAC73188.1"
/db_xref="GI:1786265"
IDQGVKQVPGAGALDIYDALTVGSDHVLVHHEQAAMHMDGLARATGEVGVVL
VTSFGATNATGATFMDISPLVLSGQVATSLIGYDAFOECMDVGISRPVVKHSF
LVKQEDIPQVKKAFWLAASRGPPVVDLPKDLINPANKLPYWPESVSMRNYPT
TGHGQIKRALQTLVAKEPVYVGGGATAGCHQQLKETVEALNLPVVGSLMGLGA
FPATHRQALGLMHGTYEANTMHNADVIFAVGVFRDRTTNLAKVCPNATVLHID
IDPISIKVTADIPVGDARQVLEQMLELISOESAHPQLDEIRWQQLQEWARQC
LKYPHSEKIKPQAVIETLRLTKGDAYVTSVSDVQHOMFAALYPPFKPRVWNSGGL
GTMGFGLPALGVKMAFPEETVVCVTGDSIQMNIQELSTALQYELPVLVNLNRRYL
GMVKQWQDMYSGRHSQSYMQSLPDFVRLAEAYGHVGIQISHPHELESKLSEALEQVR
NNRLVFVDVTVDGSEHVYPMQIRGGMDMMLSKTERT"
1998..2012
/Note="ilvI"
/Note="central position to predicted promoter:156.5"
/bound_moiety="Lrp predicted site"
2017..2045
/Note="ilvI"
/Note="factor Sigma70; promoter ilvIHp1; documented+1 at
85597"
3812..4303
/Note="ilvH"
/Note="synonym: b0078"
3812..4303
/Note="ilvH"
/EC_number="4.1.3.18"

protein_bind
/Note="central position to predicted promoter:156.5"
/bound_moiety="Lrp predicted site"
2017..2045
/Note="ilvI"
/Note="factor Sigma70; promoter ilvIHp1; documented+1 at
85597"
3812..4303
/Note="ilvH"
/Note="synonym: b0078"
3812..4303
/Note="ilvH"
/EC_number="4.1.3.18"

gene
/Note="central position to predicted promoter:156.5"
/bound_moiety="Lrp predicted site"
2017..2045
/Note="ilvI"
/Note="factor Sigma70; promoter ilvIHp1; documented+1 at
85597"
3812..4303
/Note="ilvH"
/Note="synonym: b0078"
3812..4303
/Note="ilvH"
/EC_number="4.1.3.18"

CDS
/Note="central position to predicted promoter:156.5"
/bound_moiety="Lrp predicted site"
2017..2045
/Note="ilvI"
/Note="factor Sigma70; promoter ilvIHp1; documented+1 at
85597"
3812..4303
/Note="ilvH"
/Note="synonym: b0078"
3812..4303
/Note="ilvH"
/EC_number="4.1.3.18"

Alignment Scores:
Pred. No.: 0.00385
Score: 14.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 3.12%

```

```

DB: 1 Gaps: 0
US-09-701-229-2 (1-448) x AE000118 (1-21757)
Qy 157 GluLeuTyValLeuGluLeuSerSerPheGlnLeuGluThr 170
|||||
Db 13998 GAACGTGACGTCGGAACACTGCGAGCTTCCAGCTGGAAC 14039
|||||

RESULT 14
LOCUS AE004310/c
DEFINITION Vibrio cholerae chromosome I, section 218 of 251 of the complete
chromosome.
ACCESSION AE004310 AE0003852
VERSION AE004310.1 GI:9656963
KEYWORDS
SOURCE Vibrio cholerae
ORGANISM Vibrio cholerae
Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
Vibrionaceae; Vibrio.
REFERENCE
1 (bases 1 to 22201)
Heidelberg, J.F., Eisen, J.A., Nelson, W.C., Clayton, R.A., Gwinn, M.L.,
Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Umayam, L.,
Gill, S.R., Nelson, K.E., Read, T.D., Tettelin, H., Richardson, D.,
Ermlaeva, M.D., Vamathevan, J., Bass, S., Qin, H., Dragol, I.,
Sellers, P., McDonald, L., Utterback, T., Fleischmann, R.D.,
Nierman, W.C. and White, O.
NIERMAN, W.C. and WHITE, O.
DNA sequence of both chromosomes of the cholera pathogen Vibrio
cholerae
JOURNAL Nature 406 (6795), 477-483 (2000)
MEDLINE 20406833
PUBMED 10952301
REFERENCE
2 (bases 1 to 22201)
Heidelberg, J.F., Eisen, J.A., Nelson, W.C., Clayton, R.A., Gwinn, M.L.,
Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Umayam, L.A.,
Gill, S.R., Nelson, K.E., Read, T.D., Tettelin, H., Richardson, D.,
Ermlaeva, M.D., Vamathevan, J., Bass, S., Qin, H., Dragol, I.,
Sellers, P., McDonald, L., Utterback, T., Fleischmann, R.D.,
Nierman, W.C., White, O., Salzberg, S.L., Smith, H.O., Colwell, R.R.,
Mekalanos, J.J., Venter, J.C. and Fraser, C.M.
Direct Submission
SUBMITTED (14-JUN-2000) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
FEATURES
source
1..22201
/organism="Vibrio cholerae"
/mol_type="genomic DNA"
/strain="N16961"
/serotype="O1"
/db_xref="taxon:666"
/Note="biotype: El Tor"
/complement(272..1081)
/gene="VC2391"
/complement(272..1081)
/gene="VC2391"
/Note="similar to GB:D10483 SP:P04036 GB:M10611 PID:145710
PID:216458; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="dihydrodipicolinate reductase"
/protein_id="AAF95534.1"
/db_xref="GI:9656964"
/transl_table="MVR1A1AGAAAGRMGRNLVKAHONPLSELGAGSRPSSLVGVVD
IGELCGIGKQIGIVLVNLEQAEQFVDIITDFAPASTLANLALCEQHGKLVIGTTGTF
TDAOROTLEQAARKPIVMAPNYSVGNLVKLEKAKVMGDCVCDLEIIEAHRHKV
DAPSTAIHGEATAHAMGNQLSDVAVYAREKIGTGRSRNIGTATFRAGDIEHPTA
MFADIGERVEITHATDRMTFANGAVKAAIWLAEQAGFTYIMDVLGUNDL"
complement(1196..1594)
/gene="VC2392"
/complement(1196..1594)
/gene="VC2392"
/Note="similar to GB:D10483 SP:P08337 GB:M20791 GB:X04831

```

```

PID:147795; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="mutator-3 protein"
/protein_id="AAF95535.1"
/db_xref="GI:9656965"
/translation="MKRIHIVAGIIFNSDQSEIFITKRPHLHKGGFWEFFPGKVEAG
ESREQAMVRELEEEIGITVTEQAFQHFDFDYTKLSLDFEMLVTFADQCPHGREGQ
GGWKIADLANRYFPEANDPVVKQVIAQFG"
gene
1580..1747
CDS
/gene="VC2393"
/length=168
/feature="VC2393"
/feature_start=1580
/feature_end=1747
/feature_note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF95536.1"
/db_xref="GI:9656966"
/translation="MDAFQGATPVFLADVNFSCIPFLKATAVLAVTSCLGKIKRVAL
ATLLFNKVKLA"
gene
complement(1744..4455)
/length=3011
CDS
/gene="VC2394"
/length=3011
/feature="VC2394"
/feature_start=1744
/feature_end=4455
/feature_note="similar to GB:D10483 SP:P10408 GB:M20791 PID:147794
PID:216512; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="preprotein translocase, SecA subunit"
/protein_id="AAF95537.1"
/db_xref="GI:9656967"
/translation="MITKLLTKVIGSRINDRTLRRLKIVKINNVEPAFLSDEELK
AKTVFRIQENGLDOLLPEAFATVREASKRVGMRFHFDVLIGGMVLHGQIAEM
RPEGKLTATLALYALNPGKVHIVTNDYLAKEDAETNRPFLFELGVTGVNIPN
MPQAKKAYQADILYGNNGRFDYLDNMAFREDRVRQARFPAVDEVDVSIILDE
ARTPLISGPAEDSDLYIRINKLIPLOKQKDESEYRGDGHFTVDKSKQVHLTE
TQGEVEELLVNGMNGEDTLYSPANISLLHVNALRAHFLFKNVDYIVTPDGEV
VIVDHTGTMRGMSDGLHOAVEKGVKTQENOTLASTFONFRLYEKLKSGMT
GTADTEAFQOIGLETVVITPNKPMVRNDMPVYRSEAEKFAAIIDIKRQYAKG
QPVLTYSIEKSELLSNALKAGIKHNVLNAKFEKAEIIVAEAGKPGAVIATNMA
GRDIDLVGSGQAKVEKLDNPTQEQIDAIAKFWQVHDQVLOAGSLHLITGERHER
RIDNLRGHSRGQAGSRRFYLSEMEDILLRIFTSDRMAALLQSGHDEGEALTESMLS
RSIEAKRQVRGENDIRKQLEDDVANDQRKVVELDRLMSADDISDMAQNRD
VLNAVMEYIIPQSLDMDWDIKGLEDRLNKQDLDLPQISDLADNKLLEYEALRERII
EQAVEYKAEQAEAPVARNFEKSVMLQTLDTLWKEHLAAMDHLRQGHLEGRYAKN
PROEYKRSFEDLLESLSKSDVITVLSKVQOOEVEERMEAOORRAQAEAAHQA
AOHASADAODESNQPMVYDERKVRNEPCPGSGSKYKQCHGQIN"
4809..5282
/gene="VC2395"
/length=474
CDS
/gene="VC2395"
/length=474
/feature="VC2395"
/feature_start=4809
/feature_end=5282
/feature_note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF95538.1"
/db_xref="GI:9656968"
/translation="NRDHRPTATDELIOASKLQIOEHAKATILLINROLQDILPKGLK
TOVRANVRGGNVLVLEASAAKMKVDYERHLILQLBQNGFHLISIEVRVNPYLK
QSKITSEDAARANPPPLSEHAHVLAIADQASDKVKRLOSRLAKANQND"
complement(5386..6303)
/length=1017
CDS
/gene="VC2396"
/length=1017
/feature="VC2396"
/feature_start=5386
/feature_end=6303
/feature_note="similar to GB:D10483 SP:P07652 PID:145848
PID:216510 PID:40864; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="UDP-3-O-3-hydroxymyristoyl N-acetylglucosamine
6-acetylase"
/protein_id="AAF95539.1"

```

Alignment Scores:

Pred. No.:	0.00391	Length:	22201
Score:	14.00	Matches:	14
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	3.12%	Indels:	0
DB:	1	Gaps:	0

```

/db_xref="GI:9656969"
/translation="MIRORTLKEIVKTTGVLSGRKVTLTLPRAAANTGIIVRTDVP
NPPVDFPADPSVTRMLCTALVNDQGVISVTEHNLAAAGMDINA1IEVDAPRIP
INDGASPPVYLLOQAGIOTLNAPKRFIRIKKPVRIEDGDKWAEFVPFNGFRMDFEIE
FNHPIEDGDDQLRDLVDFSSQGVKEISRRARTFGMRDIEYLQSNLCUGSDFDCAVL
DDYRLINDEGLRDNFETHKVLDAIGDLYMAGHAIVGFRAYKSGHGLNNOLLRAVL
ADQEWENATFEVEEGSPVAFAPNMVLA"
complement(6401..7597)
/gene="VC2397"
/length=1197
CDS
/gene="VC2397"
/length=1197
/feature="VC2397"
/feature_start=6401
/feature_end=7597
/feature_note="similar to GB:D10483 SP:P06138 GB:X02821 PID:216509
PID:40863; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="cell division protein FtsZ"
/protein_id="AAF95540.1"
/db_xref="GI:9656970"
/translation="MFEPMEMSDDAVIKVVGGGGGNAVEMHVMRESIEGVFEMFIN
TQAALRKTSVGTVIQGNITKGLGAGANQVGRDAALAEKRIKEFTLGADMFIA
AGMGGTGTGAAPVIAEVAKELGILTVAVTKPSPFEGKKRLAFAPQTEELSKEHVS
LITIPNEKLLKVLGRGILTEAFASANNVLKNAVQIGIELITRPGMINVDEADRVTM
SEMGHMMGSGVARGEDRAEAEAAISSPLEDIDLARGVLVNIITAGLMDRLDEF
ETVGNVTKAFASDNATVIGTSLDDPMADEIRVTVATGIGNEKKPDIITLVAGCKSKY
APVSAPVTSRVEEKAAQPLHERVEVKTPAAPAASSHSSASQSVAPKPKESGYLDIPAF
LRRQAD"
complement(7628..8890)
/gene="VC2398"
/length=1262
CDS
/gene="VC2398"
/length=1262
/feature="VC2398"
/feature_start=7628
/feature_end=8890
/feature_note="similar to GB:D10483 SP:P06137 GB:X02668 GB:M36531
GB:X02821; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="cell division protein FtsA"
/protein_id="AAF95541.1"
/db_xref="GI:9656971"
/translation="MTKTVDDNIIIVGLDIGTATVSALVGEILPDGINIAGSSPSR
GMDKGVNDESVMKSVQRAIDQAEAMAECKISHVFLSISGRHIAIRKMGMTISDE
EVSQEDMDRAIHTAKSIKIGEQRIHLVPOFTIDYQGIKNPLGSLVGMVSVHL
ITCHNDMANRIKAVERCGLKVEOLVYSLGASNAVITEDERELGVCVVDIGAGTMDI
AWTGALPHTVEFVSAGNAVTSIDIAFAGTGPVSDAEELKVKYGCALSILVSKDDTVN
VPSVGRPSRSLQRTLSEVIEPRYTELGLVNTQVDSVQESLRKKGKIKHHLAAGVVL
TGGAQIEGLVACARFVRNVRVQKPLEVGLTVYKPEYHSTAVGLLHLLAARDSQLS
DDTEYQPRPSMTGWIRLNWIKQEF"
complement(8883..9665)
/gene="VC2399"
/length=782
CDS
/gene="VC2399"
/length=782
/feature="VC2399"
/feature_start=8883
/feature_end=9665
/feature_note="similar to GB:D10483 SP:P06136 GB:X02668 GB:X02821
PID:145725; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="cell division protein FtsQ"
/protein_id="AAF95542.1"
/db_xref="GI:9656972"
/translation="MINKVLEEQRTSRPQVKHACGASFFLVLLIGGLYSTIS
WMDRELPLSLVQLQDHLHYVSALDQVRVRLDHDIGTFMSQDINVLOESVQSIIPW
SHASIRKOMPDIKTVLTVQVQVFAVNLNLDKNGTVFGDIARVNGEVKLYGPDG
TAPVQLKANRDYNPFAQLGLNLISSVLNDRRAWQIILNGLIRLEKGSLEERISRF
FLIYKQLGNKASQSVYIDRTYDGAAGVWFPQELTQERND"
complement(9847..11367)
/gene="VC2400"
/length=1520
CDS
/gene="VC2400"
/length=1520
/feature="VC2400"
/feature_start=9847
/feature_end=11367
/feature_note="similar to GB:D10483 SP:P06136 GB:X02668 GB:X02821
PID:145725; identified by sequence similarity; putative"

```



US-09-701-229-2 (1-448) x AE004310 (1-22201)

OY 157 GluLeuTyValLeuGluLeuSerPheGlnLeuGluThr 170  
|||||  
Db 14439 GAACGTGTATGCTTGAGCTATCAAGCTTCCAATTAGAACT 14398

RESULT 15

AE008699

LOCUS

DEFINITION

Salmonella typhimurium LT2, section 7 of 220 of the complete

genome.

AE008699 AE006458

VERSION

AE008699.1 GI:16418608

KEYWORDS

SOURCE

ORGANISM

Salmonella typhimurium LT2

Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

Enterobacteriaceae; Salmonella.

1 (bases 1 to 22348)

McClelland, M., Sanderson, K.E., Spieth, J., Clifton, S.W.,

Hou, S., Layman, D., Courtney, L., Porwollik, S., Ali, J., Dante, M., Du, F.,

Grewal, N., Mulvaney, E., Ryan, E., Sun, H., Florea, L., Miller, W.,

Stoneking, T., Nhan, M., Waterston, R., and Wilson, R.K.

Complete genome sequence of *Salmonella enterica* serovar Typhimurium

LT2

Nature 413 (6858), 852-856 (2001)

21534948

PUBMED

11677609

REFERENCE

2 (bases 1 to 22348)

The Salmonella typhimurium Genome Sequencing Project

Direct Submission

Submitted (29-MAR-2001) Genome Sequencing Center, Department of

Genetics, Washington University School of Medicine, 4444 Forest

Park Boulevard, St. Louis, MO 63108, USA

COMMENT Supported by NIH grant 5U 01 AI43283

Coding sequences below are predicted from manually evaluated

computer analysis, using similarity information and the programs;

GLIMMER; <http://www.tigr.org/softlab/glimmer/glimmer.html> and

GeneMark; <http://opal.biology.gatech.edu/GeneMark/>

EC numbers were kindly provided by Junko Yaburaki and the Kyoto

Encyclopedia of Genes and Genomes; <http://www.genome.ad.jp/kegg/>,

and Pedro Romero and Peter Karp at EcoCyc;

<http://ecocyc.PangeaSystems.com/ecocyc/>

The analyses of ribosome binding sites and promoter binding sites

were kindly provided by Heladia Salgado, Julio Collado-Vides and

ReguonDB;

[http://kinich.cifn.unam.mx:8850/db/regulondb\\_intro.frameset](http://kinich.cifn.unam.mx:8850/db/regulondb_intro.frameset)

This sequence was finished as follows unless otherwise noted: all

regions were double stranded, sequenced with an alternate

chemistries or covered by high quality data (i.e., phred quality >=

30); an attempt was made to resolve all sequencing problems, such

as compressions and repeats; all regions were covered by sequence

from more than one m13 subclone.

Location/Qualifiers

1..22348

/organism="Salmonella typhimurium LT2"

/mol\_type="genomic DNA"

/strain="LT2; SGSC 1412; ATCC 700720"

/db\_xref="ATCC:700720"

/db\_xref="taxon:99287"

/note="LT2"

complement(298..915)

/gene="leuB"

/note="synonym: STM0110"

complement(298..903)

/gene="leuB"

/EC\_number="4.2.1.33"

gene

CDS

/note="3-isopropylmalate dehydratase small subunit."  
(SW:LEU3\_SALTY)"  
/codon\_start=1  
/transl\_table=11  
/product="3-isopropylmalate isomerase (dehydratase),  
subunit with leuc"  
/protein\_id="AAL19074.1"  
/db\_xref="GI:16418609"  
/translation="MAKFTQFTGLVLPVLDAAVNDTDAIIPKQFLQKVRTGTGFGAHLF  
NDWFLDEKGOQNPPEVFNPEYFGASILLARENFGCGSSREHAPWALTDYGRKVI  
APSFADIFYGNSFNQLLPVTLSDAQVDELFAVKANPGIKFEVDLEAQVYKAGDKTY  
SFKIDFRHCLMGLSDIGTLQHEDAIAAYENKQAFMR"  
complement(910..915)  
/gene="leuB"  
/note="putative RBS for leuB; RegulonDB:STMS1H000490"  
complement(914..2327)  
/gene="leuB"  
/note="synonym: STM0111"  
complement(914..2314)  
/gene="leuB"  
/EC\_number="4.2.1.33"  
/note="3-isopropylmalate dehydratase large subunit."  
(SW:LEU2\_SALTY)"  
/codon\_start=1  
/transl\_table=11  
/product="3-isopropylmalate isomerase (dehydratase),  
subunit with leuc"  
/protein\_id="AAL19075.1"  
/db\_xref="GI:16418610"  
/translation="MAKTLYEKLFDAHVFEAPNETPLLYIDRLHVEVTSQAFDGL  
RAHRVPQKFTATMDHNVSTQKDNASGEMARIQMQLIKNCFEGVELYDLNH  
PYQGIHVMPQEQVTLPGMTIVCGSDTATGAFGALAFGTSEVHVLATQTLKQ  
GRATMKIEVTGNAAPGITAIDIVLAIGKTGSAGTGHVFCGDAIRALSMGRMT  
LCNMAIEMGAKAGLVAPDETTFNVKGRHAPKGRDFDEAVVEYKTLTKDDGATFDTV  
VALRAEIAQVWTGTPNGOVISVTDIIPDAPSPDVERASAKALAYNGLOPGVPL  
TDVALDRVFGSCTNSRIEDLRAAEVAKGRKVPQVQALVPGSGKAPVKAQAEGLD  
KIFTEAGFEWRLPGCSMCLAMNDRLPAGERCASISNRNFEGRGRGTHLVSPAMA  
AAAVTGHFAIRSIK"  
complement(2317..3422)  
/gene="leuB"  
/note="synonym: STM0112"  
complement(2317..3408)  
/gene="leuB"  
/EC\_number="1.1.1.85"  
/note="3-isopropylmalate dehydrogenase. (SW:LEU3\_SALTY)"  
/codon\_start=1  
/transl\_table=11  
/product="3-isopropylmalate dehydrogenase"  
/protein\_id="AAL19076.1"  
/db\_xref="GI:16418611"  
/translation="MSKNYHIAVLPGDIGEPEVMAQKAVMDAVRSRFRMTTSHYD  
VGGIADNHGHPKATVEGCEQADAILFGSVGPKWENLPPEQSPERGALLPLRKHF  
KLFSLNPAKLYOGLEAFCLRADIAANGFDILCVRELTGTVGQPKRGSGQYKEK  
AFTEVYHFEIERIAFESAKRRKRVTSIDKANVLOSSILWREIVNDVAKTYPD  
VELAHYIDNATMQLKNDPSQFVLLCSNFGDILSDCAITGSMGLPSASINEQG  
FGLYEPAGSGAPDAGNIANPQIUSLALLRLSLDCAITAIEQAINRALEEGV  
RTGDLARGAAVSTDEMDGIIRYVAEGV"  
complement(2322..2327)  
/gene="leuB"  
/note="putative RBS for leuB; RegulonDB:STMS1H000491"  
complement(3408..4999)  
/gene="leuA"  
/note="synonym: STM0113"  
complement(3408..4979)  
/gene="leuA"  
/EC\_number="4.1.3.12"  
/note="similar to E. coli 2-isopropylmalate synthase  
(AAC73185.1); Blastp hit to AAC73185.1 (523 aa), 92%

identity in aa 1 - 523"  
/codon\_start=1  
/transl\_table=11  
/product="2-isopropylmalate synthase"  
/protein\_id="AAL19077.1"  
gene  
CDS



```
Qy 157 GluLeuTyrrValLeuGluLeuSerSerPheGluLeuGluThr 170
      |||||
Db 19097 GAATTGTACGTGCTGGAATTATCCAGTTTCCAACTGGAAACT 19138
```

US-09-701-229-2 (1-448) x AE008699 (1-22348)

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus.p2n model

Run on: August 14, 2003, 04:47:05 ; Search time 351 Seconds  
(without alignments)  
3445.437 Million cell updates/sec

Title: US-09-701-229-2

Perfect score: 2260  
Sequence: 1 MSLIASDHFRIVVGLKSGM.....MFKNFERGRLEPAKAVEELA 448

Scoring table:  
BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 3294098

Minimum DB seq length: 25  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+p2n.model -DEV=xlp  
-Q=/cgn2\_1/USPTO.spool/US09701229/runat\_09082003\_161156\_4371/app\_query.fasta\_1.647  
-DB=N\_Geneseq\_19Jun03 -Qfmt=fastp -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPEL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS-human40.cdi  
-LIST=45 -DOCALIGN=200 -THR\_SCORE=pcpt -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTPMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=25 -MAXLEN=2000000000  
-USPR=US09701229.ecgn.1.1.471 -runat\_09082003\_161156\_4371 -NCPU=6 -ICPU=3  
-NO\_WMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N\_Geneseq\_19Jun03:  
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:  
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:  
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:  
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:  
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:  
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:  
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:  
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:  
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:  
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:  
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:  
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:  
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:  
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:  
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:  
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:  
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:  
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:  
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:  
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:  
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:  
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:  
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:  
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:  
25: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2003.DAT:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2260	100.0	1347	23	Pseudomonas aerugi
2	2260	100.0	1450	21	Pseudomonas aerugi
3	989.5	43.8	1317	23	Salmonella typhi D
4	985.5	43.6	1317	22	Escherichia coli p
5	985.5	43.6	1317	23	E. coli DNA for ce
6	889.5	39.4	96109	22	Genomic fragment #
7	879	38.9	1314	23	Haemophilus influe
8	879	38.9	1830121	17	Haemophilus influe
9	716	31.7	640681	24	Buchnera sp. genom
10	704	31.2	1380	21	Neisseria meningit
11	704	31.2	69936	21	N. meningitidis pa
12	704	31.2	349980	21	Neisseria meningit
13	704	31.2	1437668	21	N. meningitidis B
14	701	31.0	1335	25	N. gonorrhoeae nuc
15	603.5	26.7	2193	20	Streptococcus pneu
16	594.5	26.3	1350	25	S. pneumoniae type
17	594.5	26.3	6693	19	Streptococcus pneu
18	594.5	26.3	2162598	25	Streptococcus pneu
19	592.5	26.2	1353	23	Streptococcus pneu
20	586.5	26.0	1350	18	Streptococcus pneu
21	586.5	26.0	1350	20	S. pneumoniae Murd
22	586.5	26.0	1353	20	S. pneumoniae Murd
23	586.5	26.0	1353	23	Streptococcus pneu
24	577.5	25.6	6217	19	DNA encoding S. pn
25	542	24.0	1080	23	DNA encoding novel
26	542	24.0	1080	23	DNA encoding novel
27	542	24.0	1080	23	DNA encoding novel
28	537.5	23.8	2703	23	DNA encoding novel
29	537.5	23.8	2703	23	DNA encoding novel
30	534.5	23.7	1352	24	Bacillus lichenifo
31	529.5	23.4	1353	24	Streptococcus poly
32	529.5	23.4	1365	24	Streptococcus poly
33	529.5	23.4	2155561	22	Streptococcus poly
34	525	23.2	1461	22	Mycobacterium tube
35	525	23.2	4411529	22	Mycobacterium tube
36	519	23.0	1368	23	Enterococcus faeca
37	519	23.0	1371	23	Enterococcus faeca
38	515	22.8	4403765	22	Mycobacterium tube
39	510	22.6	4942	23	DNA encoding novel
40	510	22.6	4944	23	DNA encoding novel
41	510	22.6	4944	23	DNA encoding novel
42	509.5	22.5	2365589	24	Genomic sequence o
43	506.5	22.4	1356	24	Streptococcus poly
44	506.5	22.4	1407	20	Streptococcus pyog
45	493	21.8	1003	19	Streptococcus pneu

ALIGNMENTS

RESULT 1  
AAS54275  
ID AAS54275 standard; DNA; 1347 BP.  
AC AAS54275;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE Pseudomonas aeruginosa DNA for cellular proliferation protein #406.  
XX  
KW Antisense; ds; prokaryotic cellular proliferation gene;  
XX  
KW antibiotic; antibacterial; drug design.  
XX  
OS Pseudomonas aeruginosa.  
XX  
PN WO200170955-A2.  
XX

PD 27-SEP-2001.  
 XX  
 PF 21-MAR-2001; 2001WO-US09180.  
 XX  
 PR 21-MAR-2000; 2000US-191078P.  
 PR 23-MAY-2000; 2000US-206948P.  
 PR 26-MAY-2000; 2000US-207727P.  
 PR 23-OCT-2000; 2000US-242578P.  
 PR 27-NOV-2000; 2000US-253625P.  
 PR 22-DEC-2000; 2000US-257931P.  
 PR 16-FEB-2001; 2001US-269308P.  
 XX  
 PA (ELIT-) ELITRA PHARM INC.  
 XX  
 PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;  
 PI Yamamoto RT, Xu HH;  
 XX  
 XX WPI: 2001-611495/70.  
 DR P-PSDB; A036416.  
 XX  
 PT New polynucleotides for the identification and development of  
 PT antibiotics, comprise sequences of antisense nucleic acids -  
 PS  
 PS Claim 27; Seq ID No 7912; 511pp; English.  
 XX  
 CC The invention relates to antisense inhibitors of genes essential to  
 CC prokaryotic cellular proliferation, their use in identifying the  
 CC genes, their use in the discovery of novel antibiotics, the essential  
 CC genes themselves and the encoded proteins. The prokaryotes used are  
 CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella  
 CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The  
 CC invention is also useful for the identification of potential new targets  
 CC for antibiotic development. The antisense nucleic acids can also be used  
 CC to identify proteins used in proliferation, to express these proteins,  
 CC and to obtain antibodies capable of binding to the expressed proteins.  
 CC The proteins can be used to screen compounds in rational drug discovery  
 CC programmes. The antisense nucleic acid sequence is also useful to screen  
 CC for homologous nucleic acids which are required for cell proliferation in  
 CC a wide variety of organisms. The present sequence encodes an  
 CC essential prokaryotic cellular proliferation protein.  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 1347 BP; 214 A; 453 C; 453 G; 227 T; 0 other;

Alignment Scores:  
 Pred. No.: 1, 27e-205 Length: 1347  
 Score: 2260.00 Matches: 448  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 23 Gaps: 0

US-09-701-229-2 (1-448) x AAS54275 (1-1347)

QY 1 MetSerLeuIleAlaSerAspHisPheArgIleValValGlyLeuGlyLysSerGlyMet 20  
 DB 1 ATGAGCCTGATCGCTCCGACCACCTTCGCGATCGTTGTCGGCTCGGCAAGAGCGCATG 60  
 QY 21 SerLeuValArgTyrLeuAlaArgArgGlyLeuProPheAlaValValAspThrArgGlu 40  
 DB 61 TCCTGTGTGGCTACCTGGCGCGCGCGCTTTCGCGCTTTCGCGCTGATACCGGAGAG 120  
 QY 41 AsnProGluLeuAlaThrLeuArgAlaGlnTyrProGlnValGluValArgCysGly 60  
 DB 121 AACCCGCGGAGCTGGCCACCCCTGCGTCCCGATATCCGAGGTGAAGTGGTTCGCGC 180  
 QY 61 GluLeuAspAlaGluPheLeuCysSerAlaArgGluLeuTyrValSerProGlyLeuSer 80  
 DB 161 GAACCTCGAGCGGAGTTCCTCTGTCTCCGCGCGGACCTCTATGTCAGCCCGCGCTTTCG 240

QY 81 LeuArgThrProAlaLeuValGlnAlaAlaAlaLysGlyValArgIleSerGlyAspIle 100  
 DB 241 CTGCGCACCCCTGCGCTGTGTACAGCGCGCGGAAAGGCGTGGCATCTCCGGTGCATC 300  
 QY 101 AspLeuPheAlaArgGluAlaLysAlaProIleValAlaIleThrGlySerAsnAlaLys 120  
 DB 301 GATCTCTTCGCGCGGAGCGGAAAGCGCGCATGCTGGCGCATACCGGGTTCGAACGGAAG 360  
 QY 121 SerThrValThrThrLeuValGlyGluMetAlaValAlaAlaAspLysArgValAlaVal 140  
 DB 361 AGCACCGTGACCACTTGTGGCGGAAATGGCGGTGGCGCGGACAGCGTGTCCGCGTC 420  
 QY 141 GlyGlyAsnLeuGlyThrProAlaLeuAspLeuLeuAlaAspIleLeuTyrVal 160  
 DB 421 GCGCGCAACCTCGGCACCGCGCGCTCGACCTGCTGGCGAGCATCGAGTGTACGTCG 480  
 QY 161 LeuGluLeuSerSerPheGlnLeuGluThrCysAspArgLeuAsnAlaGluValAlaThr 180  
 DB 481 TTGGAGCTGTGAGCTTCCAGCTGGAACCTCGGATCGCTCAAGCCGAGGTGGGACCC 540  
 QY 181 ValLeuAsnValSerGluAspHisMetAspArgTyrAspGlyMetAlaAspTyrHisLeu 200  
 DB 541 GTGCTGAACGTCAGCGAAGACCATATGGATCGCTACGACGGCATGGCTACACCTG 600  
 QY 201 AlaLysHisArgIlePheArgGlyAlaArgGlnValValValAsnArgAlaAspAlaLeu 220  
 DB 601 GCCAAGCACCGGATCTTCGCGGTCCGCGCAGGTGCTGGTGAATCGCGCCGATGCCCTG 660  
 QY 221 ThrArgProLeuIleAlaAspThrValProCysTyrSerPheGlyLeuAsnLysProAsp 240  
 DB 661 ACCCGACCGTGATCGCGATACCGTGGCTGCTGGTGGTGGCTGACCAAGCGCGGAC 720  
 QY 241 PheLysAlaPheGlyLeuIleGluAspGlyGlnLysTrpLeuAlaPheGlnPheAsp 260  
 DB 721 TTCAAGGCTTTCGCGCTGATCGAGGAAGACGCCAGAGTGGCTGGCTTCAGATTCGAC 780  
 QY 261 LysLeuLeuProValGlyLeuLysIleArgGlyAlaHisAsnTyrSerAsnAlaLeu 280  
 DB 781 AAGTGTCTGCGGTTGGCGAATGAAGATCCGTCGCGCCCACTATTTCCAAACGCGTC 840  
 QY 281 AlaAlaLeuAlaLeuGlyHisAlaValGlyLeuProPheAspAlaMetLeuGlyAlaLeu 300  
 DB 841 GCGCGCTGGCGTGGCCATCGCGTGGCTGCGCTTCGAGCCCATGCTCGCGCGCTG 900  
 QY 301 LysAlaPheSerGlyLeuAlaHisArgCysGlnTrpValArgGluArgGlnGlyValSer 320  
 DB 901 AAGCGCTTTCGCGCTGCTCATCGCTGCCAGTGGGTACGCGAGCGGCGGCGCGTGCAGC 960  
 QY 321 TyrTyrAspAspSerLysAlaThrAsnValGlyAlaAlaLeuAlaAlaIleGluGlyLeu 340  
 DB 961 TACTACGACGATTCGAAGGCCACCAACGTCGCGCGCGCGCTTCGCGCGCATCGAGGGCTG 1020  
 QY 341 GlyAlaAspIleAspGlyLysLeuValLeuLeuAlaGlyLysGlyValAspGlyAlaAsp 360  
 DB 1021 GTGCGGACATCGACGCGCAAGCTGTGTGCTGCGCGCGGAGAGAGCGGCGCGCGAT 1080  
 QY 361 PheHisAspLeuArgGluProValAlaAlaArgPheCysArgAlaValValLeuLeuGlyArg 380  
 DB 1081 TTCATGACCTGCGCGAGCGGTGCGCGCTTCTCCGCGCGGTGTGTACTGTTGCCCT 1140  
 QY 381 AspAlaGlyLeuIleAlaGlnAlaLeuGlyAsnAlaValProLeuValArgValAlaThr 400  
 DB 1141 GACGCGGCGCTGATTGCGCGACCTGGGCGACGCGGTACCGCTGGTGGCTGCGCTGCAAGC 1200  
 QY 401 LeuAspGluAlaValArgGlnAlaGluLeuAlaArgGluGlyAspAlaValLeuLeu 420  
 DB 1201 CTGGACGAAGCAGTCCGCGAGCGCCCGAGCTGGCGCGCGGAGAGCGGCGGTGTGTG 1260  
 QY 421 SerProAlaCysAlaSerLeuAspMetPheLysAsnPheGluGluArgGlyArgLeuPhe 440  
 DB 1261 TCGCGCGCTCGCGAGCGCTGGACATGTTCAAGAACTTCGAAGAACGCGGCGCTGTTC 1320  
 QY 441 AlaLysAlaValGluGluLeuAla 448

```

Db      1321  GCCAAAGCCGTAGAGGAGTAGCG 1344
RESULT 2
AAZ47131
ID      AAZ47131 standard; DNA; 1450 BP.
XX      AC
XX      AAZ47131;
XX      28-MAR-2000 (first entry)
XX      Pseudomonas aeruginosa murD gene.
DE      MurD; bacterial cell wall; biosynthesis; inhibitor; D-glutamate;
KW      UDP-N-acetylmuramyl-L-alanine precursor; ds.
XX      Pseudomonas aeruginosa.
OS      W09961050-Al.
XX      02-DEC-1999.
XX      26-MAY-1999; 99WO-US11585.
XX      29-MAY-1998; 98US-0087308.
XX      (MERI ) MERCK & CO INC.
XX      El-Sherbelini M, Azzolina B;
XX      WPI; 2000-072548/06.
DR      P-PSDB; AAY56047.
XX      New nucleic acid encoding the MurD protein of Pseudomonas aeruginosa,
PT      used to identify specific inhibitors -
XX      Claim 4; Page 8-9; 35pp; English.
XX      This sequence represents the MurD gene from Pseudomonas aeruginosa.
CC      MurD is an enzyme involved in biosynthesis of bacterial cell walls,
CC      so inhibiting it inhibits bacterial growth (by preventing
CC      MurD-catalyzed addition of D-glutamate to an alanyl residue in the
CC      UDP-N-acetylmuramyl-L-alanine precursor). Cells that express murD
CC      protein can be used: (a) to identify inhibitors of MurD, and (b) for
CC      recombinant production of MurD, which is used to raise specific
CC      antibodies (potential therapeutic inhibitors). Fragments of the murD
CC      gene can be used as probes and primers to detect murD-encoding sequences
CC      (to identify polymorphic murD genes or to detect P. aeruginosa), also
CC      as antisense modulators. The murD inhibitors are useful for treating or
CC      preventing bacterial infections, including those caused by P. aeruginosa.
XX      Sequence 1450 BP; 227 A; 487 C; 487 G; 249 T; 0 other;

Alignment Scores:
Pred. No.:      1.4e-205      Length:      1450
Score:          2260.00      Matches:      448
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:    100.00%      Indels:      0
DB:            21      Gaps:        0

US-09-701-229-2 (1-448) x AAZ47131 (1-1450)
Qy      1 MetSerLeuIleAlaSerAspHisPheArgTleValValGlyLeuGlyLysSerGlyMet 20
Db      51 ATGAGCCTGATCGCTCCGACCACCTCCGGATCGTGTGGCCCTCGGCAAGCGGCATG 110
Qy      21 SerLeuValArgTyrLeuAlaArgArgGlyLeuProPheAlaValAlaAspThrArgGlu 40
Db      111 TCCCTGTGCGCTACCTGGCGCGCGCGCTTTCGCGGTGTCGATACCCGAGAG 170
Qy      41 AsnProProGluLeuAlaThrLeuArgAlaGlnTyrProGlnValGluValArgCysGly 60

```

```

171 AACCCGCGGAGCTGGCCACCCTCGTGGCCAGTATCCGAGTGGAAAGTGCCTGGCGC 230
Qy      GluLeuAspAlaGluPheLeuCysSerAlaArgGluLeuTyrValSerProGlyLeuSer 80
Db      231 GAACCTGACGCCGAGTTCTCTGCTCCGCCCGGAACTCTATGTACGCCCGGCTGTGCG 290
Qy      81 LeuArgThrProAlaLeuValGlnAlaAlaLysGlyValArgGlySerGlyAspIle 100
Db      291 CTGCGCACCCCTGCTGGTACAGCCCGCGAAAGCGGTGCGCATCTCCGGTGACATC 350
Qy      101 AspLeuPheAlaArgGluAlaLysAlaProIleValAlaIleThrGlySerAsnAlaLys 120
Db      351 GATCTCTCGCCCGAGGCGGAGCCCGCATCGTCGCCATCAGCGGTTCACACGCGAAG 410
Qy      121 SerThrValThrThrLeuValGlyGluMetAlaValAlaAlaAspLysArgValAlaVal 140
Db      411 AGCACCGTGACCCCTGGTGGCGAAATGGCGGTGGCCGGACACGCGTGTCCCGCTC 470
Qy      141 GlyGlyAsnLeuGlyThrProAlaLeuAspLeuLeuAlaAspAspIleGluLeuTyrVal 160
Db      471 GCGCGCAACCTCGGCACCCCGCGCTCGACCTGCTGGCCGACGACATCGAGCTGTACGTG 530
Qy      161 LeuGluLeuSerSerPheGlnLeuGluThrCysAspArgLeuAsnAlaGluValAlaThr 180
Db      531 TTGGAGCTGTGAGCTTCCAGCTGGAACCTGCGATCGCTCAACGCGGAGGTGGCGACC 590
Qy      181 ValLeuAsnValSerGluAspHisMetAspArgTyrAspGlyMetAlaAspTyrHisLeu 200
Db      591 GTGCTGAACGTCAGCGAGACCATATGATCGTACGACGCGATGGCTGACTACCCACCTG 650
Qy      201 AlaLysHisArgIlePheArgGlyAlaArgGlnValValAlaAsnArgAlaAspAlaLeu 220
Db      651 GCCAAGCACCGGATCTTCCGCGGTGCCCGCAGCTCGTGGTGAATCGCGCGCATGTCCTG 710
Qy      221 ThrArgProLeuIleAlaAspThrValProCysTrpSerPheGlyLeuAsnLysProAsp 240
Db      711 ACCCGACCGCTGATCGCGCATACCGTGGCTGGTGGTGGTGGTGGTGGTGGTGGTGGT 770
Qy      241 PheLysAlaPheGlyLeuIleGluGluAspGlyGlnLysTrpLeuAlaPheGlnPheAsp 260
Db      771 TTCAAGGCTTTCGGCTGATCGAGAAAGACGCCAGAGTGGCTGGCTTCCAGTTCGAC 830
Qy      261 LysLeuLeuProValGlyGluLeuLysIleArgGlyAlaHisAsnTyrSerAsnAlaLeu 280
Db      831 AAGTGTGTCGCGTGGCGAACTGAAGATCGTGGCGGCCACAACTATTCCAACGCGCTC 890
Qy      281 AlaAlaLeuAlaLeuGlyHisAlaValGlyLeuProPheAspAlaMetLeuGlyAlaLeu 300
Db      891 GCCCGCTGGCGCTGGCGCATCGGTCGCGCTGGCTGGTGGTGGTGGTGGTGGTGGTGGT 950
Qy      301 LysAlaPheSerGlyLeuAlaHisArgCysGlnTrpValArgGluArgGlnGlyValSer 320
Db      951 AAGCGGTTTCCGGCTGGCTCATCGTGGCAGTGGGTACCGGAGCGCGAGCGGCGGTGAGC 1010
Qy      321 TyrTyrAspSerLysAlaThrAsnValGlyAlaAlaLeuAlaAlaIleGluGlyLeu 340
Db      1011 TACTACGACGATTCCAAGGCCACCAACGTCGCGCGCGCCCTCGCGCGCATCGAGGGCTG 1070
Qy      341 GlyAlaAspIleAspGlyLysLeuValLeuAlaGlyArgGlyAspGlyLysGlyAlaAsp 360
Db      1071 GGTGCCGACATCGAGCGCAAGCTGGTGTCTCGCCGCGGAGACGCGCAAGGGCGCCCAT 1130
Qy      361 PheHisAspLeuArgGluProValAlaArgPheCysArgAlaValValLeuLeuGlyArg 380
Db      1131 TTCCATGACCTGCGCGAGCGCGCTCGCGCTTCTGCGCGCGGTGGTACTGCTGGCGCT 1190
Qy      381 AspAlaGlyLeuIleAlaGlnAlaLeuGlyAsnAlaValProLeuValArgValAlaThr 400
Db      1191 GACCGCGGCTGATTGCCAGGCACCTGGGCAACGCGGTACCGCTGGTGGCTGGCTGGCAACG 1250
Qy      401 LeuAspGluAlaValArgGlnAlaAlaGluLeuAlaArgGluGlyAspAlaValLeuLeu 420
Db      1251 CTGGACGAAGCAGTCCGCGAGCGCGGAGCTGGCGCGGAGCGATGCGGTGCTGTGTTG 1310

```

QY 421 SerProAlaCysAlaSerLeuaspMetPheLysAsnPheGluGluArgGlyArgLeuPhe 440  
 |||||  
 Db 1311 TCGCCGCGCTGGCGAGCTGGACATGTTCAAGAACTTCAAGAACGCGGAGCGCTGTTC 1370  
 |||||  
 QY 441 AlaLysAlaValGluGluLeuAla 448  
 |||||  
 Db 1371 GCCAAGCGGTAGAGGAGCTAGCG 1394

## RESULT 3

AAS56341

ID AAS56341 standard; DNA; 1317 BP.

XX AC AAS56341;

XX DT 13-FEB-2002 (first entry)

XX DE Salmonella typhi DNA for cellular proliferation protein #374.

XX KW Antisense; ds; prokaryotic cellular proliferation gene;

XX KW antibiotic; antibacterial; drug design.

XX OS Salmonella typhi.

XX PN WO200170955-A2.

XX PD 27-SEP-2001.

XX PF 21-MAR-2001; 2001WO-US09180.

XX PR 21-MAR-2000; 2000US-191078P.

XX PR 23-MAY-2000; 2000US-206848P.

XX PR 26-MAY-2000; 2000US-207272P.

XX PR 23-OCT-2000; 2000US-242578P.

XX PR 27-NOV-2000; 2000US-253625P.

XX PR 22-DEC-2000; 2000US-257931P.

XX PR 16-FEB-2001; 2001US-269308P.

XX PA (ELIT-) ELITRA PHARM INC.

XX PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;

XX PI Yamamoto RT, Xu HH;

XX WPI: 2001-611495/70.

XX P-PSDB: AAU38482.

XX PT New polynucleotides for the identification and development of

XX PT antibiotics, comprise sequences of antisense nucleic acids -

XX PS Claim 27; Seq ID No 9978; 51lpp: English.

XX CC The invention relates to antisense inhibitors of genes essential to  
 CC prokaryotic cellular proliferation, their use in identifying the  
 CC genes, their use in the discovery of novel antibiotics, the essential  
 CC genes themselves and the encoded proteins. The prokaryotes used are  
 CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella  
 CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The  
 CC invention is also useful for the identification of potential new targets  
 CC for antibiotic development. The antisense nucleic acids can also be used  
 CC to identify proteins used in proliferation, to express these proteins,  
 CC and to obtain antibodies capable of binding to the expressed proteins.  
 CC The proteins can be used to screen compounds in rational drug discovery  
 CC programmes. The antisense nucleic acid sequence is also useful to screen  
 CC for homologous nucleic acids which are required for cell proliferation in  
 CC a wide variety of organisms. The present sequence encodes an  
 CC essential prokaryotic cellular proliferation protein.

XX CC Note: The sequence data for this patent did not form part

XX CC of the printed specification, but was obtained in electronic

XX CC format directly from WIPO at

XX CC ftp.wipo.int/pub/published\_pct\_sequences.

XX XX Sequence 1317 BP; 258 A; 341 C; 427 G; 291 T; 0 other;

## Alignment Scores:

Pred. No.:

Score: 1.24e-84

Percent Similarity: 989.50

Best Local Similarity: 64.56%

Query Match: 51.92%

DB: 43.78%

Length: 1317

Matches: 230

Conservative: 56

Indels: 21

Gaps: 9

US-09-701-229-2 (1-448) x AAS56341 (1-1317)

QY 11 IleValValGlyLeuLysSerGlyMetSerLeuValArgTyrLeuAlaArgArgGly 30  
 |||||  
 Db 28 GTCATTATCGGTCTGGGCTTACCGGACTCTCTGTGGTGACTTTTTCGCCCGCGGC 87  
 |||||  
 QY 31 LeuProPheAlaValValAspThrArgGluAsnProGluLeuAlaThrLeuArgAla 50  
 |||||  
 Db 88 GTGACGCCGCGGTGATGATGATCTCGCTGACGCCGCGGTCTGGTAAGCTG----- 141  
 |||||  
 QY 51 GlnTyrProGln---ValGluValArgCysGlyLeuAspAlaGluPheLeuCysSer 69  
 |||||  
 Db 142 -----CCGCAAGAGGTTGAGCGTCAAGTTGGCGGCTGAACGACGAGTGCTTACGC 195  
 |||||  
 QY 70 AlaArgGluLeuTyrValSerProGlyLeuSerLeuArgThrProAlaLeuGlnAla 89  
 |||||  
 Db 196 GCGGATTAAATCGTCCGACGCCCTGGTATTGCTGCGGCATCCCTCTCTTAGCGCTGCC 255  
 |||||  
 QY 90 AlaAlaLysGlyValArgIleSerGlyAspIleAspLeuPheAlaArgGluAlaLysAla 109  
 |||||  
 Db 256 GCCAGCGCTGGAGTGGAAATTTGCGCGCATATCGAACTGTTTGGCGCGAAGCGAAGCG 315  
 |||||  
 QY 110 ProIleValAlaIleThrGlySerAsnAlaLysSerThrValThrThrLeuValGlyGlu 129  
 |||||  
 Db 316 CCGATTGTGGCCATCACCGGCTCGAAGCAAAAGCACCGCTGACCACTTAGTGGCGGAG 375  
 |||||  
 QY 130 MetAlaValAlaAlaAspLysArgValAlaValAlaGlyGlyAsnLeuGlyThrProAlaLeu 149  
 |||||  
 Db 376 ATGGCGAAGCGCGGCGCTCAATGTCGCGGTGGCGGCAATATCGTCTCCGCGCGTGG 435  
 |||||  
 QY 150 AspLeuLeuAlaAspAspIleGluLeuTyrValLeuGluLeuSerSerPheGlnLeuGlu 169  
 |||||  
 Db 436 ATGCTGCTGGATGCGCGATCGCGAATTTGTACGTGTGGAATTTATCCAGTTTCCAAATGGAA 495  
 |||||  
 QY 170 ThrCysAspArgLeuAsnAlaGluValAlaThrValLeuAsnValSerGluAspHisMet 189  
 |||||  
 Db 496 ACTACCTCAAGTTTTCAGCGCGCGCGGCAACGCTGCTCAAGCTCACTGAAGATCATATG 555  
 |||||  
 QY 190 AspArgTyrAsp---GlyMetAlaAspTyrHisLeuAlaLysHisArgIlePheArgGly 208  
 |||||  
 Db 556 GACCGGTATCCGTTGGTTTGGCAACAGTACCGCGGCGGCAAACTGCGCGCTCTACGAGAAG 615  
 |||||  
 QY 209 AlaArgGlnValValValAsnArgAlaAspAlaLeuThrArgProLeuIleAlaAspThr 228  
 |||||  
 Db 616 GCGAAAGTGTGGTAGTGAATCCCGATGACGCGTGTGACGATCGCGTACGCGGCGCGAT 675  
 |||||  
 QY 229 ValProCysTrpSerPheGlyLeuAsnLysProAspPheLysAlaPheGlyLeuIleGlu 248  
 |||||  
 Db 676 GAGCGCTGCTGAGCTTTGGCGTCAATATGGGTGATTATCAC-----CTTAATCGT 726  
 |||||  
 QY 249 GluAspGlyGlnLysTrpLeuAlaPheGlnPheAspLysLeuLeuProValGlyGluLeu 268  
 |||||  
 Db 727 CAGCAGGCGGAACCTGGCTAGCTCAAGGTGAGAGGTGCTCAACGTGAAGAGATG 786  
 |||||  
 QY 269 LysIleArgGlyAlaHisAsnTyrSerAsnAlaLeuAlaLeuAlaLeuGlyHisAla 288  
 |||||  
 Db 787 AAGCTTTCGGTCAAGTAACTATACCAATGTTAGCGGCGCTGCGCGTGGCGATGCC 846  
 |||||  
 QY 289 ValGlyLeuProPheAspAlaMetLeuGlyAlaLeuLysAlaPheSerGlyLeuAlaHis 308  
 |||||  
 Db 847 GTAGTCTGCCGCGCGCGAGTGTGAAGCGTGTGACGACATTTACCGGTCTGGCGCAT 906  
 |||||  
 QY 309 ArgCysGlnTrpValArgGluArgGlnGlyValSerTyrTyrAspSerLysAlaThr 328  
 |||||  
 Db 907 CGCTTCCAGCTGGCGTTGGAGCATAAACGCGCTTCGTTGGATCAACGACTCGAAGACGACC 966  
 |||||

Qy	329	AsnValGlyAlaAlaLeuAlaLalleGluGlyLeuGlyAlaAspIleAspGlyLysLeu	348
		:::	:::
Db	967	AATGTCCGCAGTACCGAAGCGCGCTAAACGGTTTG-----CATGTGGACGGTAGCCTG	1020
Qy	349	ValLeuLeuAlaGlyGlyaspGlyLysGlyAlaAspPheHisAspLeuArgGluProVal	368
Db	1021	CATCTGCTGTGGCGCGCACGGTAAGTCGGCAGACTTT-----TCTCGGCTG	1068
Qy	369	AlaArgPhecysArgAla-----ValValLeuLeuGlyArgAspAlaGlyLeu	384
		:::	
Db	1069	GCGCGGTATTGTACC GGCGATCGTATCCGCTGTATTGCTTTGGGCGCGATGCGCGCAG	1128
Qy	385	IleAlaGlnAlaLeuGlyAsnAlaValProLeuValArgValAlaThrLeuAspGluAla	404
		:::                :::	
Db	1129	CTTGCC---GCATCTGGTCCGGGAATC---GCCCAACAGACTGAGAGTGGGAAGAGCG	1182
Qy	405	ValArgGlnAlaAlaGluLeuAlaArgGluGlyAspAlaValLeuLeuSerProAlaCys	424
		:::                :::	
Db	1183	ATCGGTTTGTGTCGCGCGCACGTTTCACCGCGGTGATATGGTGTCTGTGTGTCGCCCGCCTGC	1242
Qy	425	AlaSerLeuAspMetPheYlsAsnPheGluGluArgLeuPheAlaLysAlaVal	444
Db	1243	GCCAGCCTCGATCAGTTTAAAAATTTTGAGCAACGGCGGATGCTTTACCCGCTGTCGC	1302
Qy	445	GluGluLeu	477
		:::	
Db	1303	AAGCAGTTA	1311

## RESULT 4

AAH81460	AAH81460 standard; DNA; 1317 BP.
XX	
XX	AAH81460;
XX	
XX	21-SEP-2001 (first entry)
DT	
XX	
XX	Escherichia coli protein encoding nucleotide sequence SEQ ID NO:259.
DE	
XX	Escherichia coli; identification; proliferation; microorganism;
KW	antimicrobial; antibacterial; antibiotic; gene therapy; diagnosis;
KW	bacterial growth inhibition; ds.
XX	
XX	
OS	Escherichia coli.
XX	
XX	WO200148209-A2.
PN	
XX	05-JUL-2001.
PD	
XX	19-DEC-2000; 2000WO-US34419.
XX	
PF	
XX	23-DEC-1999; 99US-0173005.
PR	
XX	(ELIT-) ELITRA PHARM INC.
XX	
XX	Forsyth RA, Ohlsen KL, Zyskind JW;
PI	
XX	
XX	WPI; 2001-457376/49.
DR	
XX	P-PSDB; AAG98404.
XX	
XX	Novel nucleic acids encoding proteins required for Escherichia coli
PT	proliferation, useful for screening for antimicrobial agents -
PT	
XX	
XX	Claim 9; Page 388-389; 596pp; English.
PS	
XX	
XX	The present invention describes a purified or isolated nucleic acid
CC	sequence (I) consisting essentially of one of the 93 nucleotide sequences
CC	given in AAH81202 to AAH81294, where expression of the nucleic acid in a
CC	microorganism is capable of inhibiting proliferation of a microorganism.
CC	(I) have antibacterial and antibiotic activities, and can be used in
CC	gene therapy. Expression of (I) in a microorganism inhibits proliferation
CC	of the microorganism, and the manufactured antibiotic is useful for
CC	reducing the activity or level of a gene product required for
CC	

proliferation of a microorganism in a subject, specifically humans. The nucleic acids that inhibit bacterial growth or proliferation can be used as antisense therapeutics for killing bacteria. In addition to therapeutic applications, the nucleic acid sequences complementary to sequences required for proliferation can be used as diagnostic tools. For example, nucleic acid probes complementary to proliferation-required sequences that are specific for particular species of microorganisms can be used as probes to identify particular microorganism species in clinical specimens. AAH81295 to AAH81487 encode the *Escherichia coli* proteins given in AAG98239 to AAG98431, and AAH81488 to AAH81491 represent oligonucleotides, which are used in the exemplification of the present invention.

Sequence 1317 BP: 294 A: 323 C: 400 G: 300 T: 0 other;

Alignment Scores:	
Pred. NO.:	2,988-84
Score:	985.50
Percent Similarity:	64.33%
Best Local Similarity:	51.24%
Query Match:	43.61%
DB:	22
Length:	1317
Matches:	227
Conservative:	58
Mismatches:	137
Indels:	21
Gaps:	9

US-09-701-229-2 (1-448) x AAH81460 (1-1317)

Qy	11	IleValValGlyLeuGlyLysSerGlyMetSerLeuValArgTyrLeuAlaIhrGly	30
Db	28	GTCAATTACGGCTGCGGCTCAACGGGCTTCTCGTGGACTTTTTCCTCGCTCGCGGT	87
Qy	31	LeuProPheAlaValValAspThrArgGluAsnProGluLeuAlaThrLeuArgAla	50
Db	88	GTACGCGCGCGCTTATGGATACGCGTATGACACCGCGCTGGCTGGATAAATTA	14
Qy	51	GlnTyrProGln---ValGluValArgCysGlyLeuAspAlaGluPheLeuCysSer	69
Db	142	-----CCCGAAGCCGTAGAACCCACACGGGCAGTCTGAATGATGAATGGCTGATGGC	19
Qy	70	AlaArgGluLeuTyrValSerProGlyLeuSerLeuArgThrProAlaLeuValGlnAla	89
Db	196	GCAGATCTGATGTCGCGCTCAACGGGTATTCGACTGGCGCATCTCAATCCCTAAACGCGTGC	25
Qy	90	AlaAlaLysGlyValArgIleSerGlyAspIleAspLeuPheAlaArgGluAlaLysAla	10
Db	256	GCTGATGCCGGAATCGAAATCGTTGGCGATATCGAGCTGTTCTGTGCGGAACGACAAGCA	31
Qy	110	ProIleValAlaIleThrGlySerAsnAlaLysSerThrValThrThrLeuValGlyGlu	12
Db	316	CCGATTGTGCGGATTACCGGTCTACGGCAAAAGACCGGTCAACACGCTAGTGGGTGAA	37
Qy	130	MetAlaValAlaAlaAspLysArgValAlaValGlyLysLeuGlyThrProAlaLeu	14
Db	376	ATGCGGAAACGCGGGGGTTAACGTTGGTGTGGTGGCAATATTGGCTCGCTCGGTG	43
Qy	150	AspLeuLeuAlaAspAspIleGluLeuTyrValLeuGluLeuSerSerPheGlnLeuGlu	16
Db	436	ATGCTACTGATGATGATGTGAACGTAGCTGCTGGAACCTGTGAGAGTTCACAGCTGGAA	49
Qy	170	ThrCysAspArgLeuAsnAlaGluValAlaIleThrValLeuAsnValSerGluAspHisMet	18
Db	496	ACCACCTCCAGCTTACAGCGGGTAGCGGACCACTTCTGAACGTGACTGAAGATCATATG	55
Qy	190	AspArgTyrAsp---GlyMetAlaAspTyrHisLeuAlaLysHisArgIlePheArgGly	20
Db	556	GATCGCTATCCGTTTGGTTTACAAACATATCGTCGACGAAAACTGCGCATATTCAGNAAC	61
Qy	209	AlaArgGlnValValValAsnArgAlaAspAlaLeuThrArgProLeuLeuAlaAspThr	22
Db	616	GCGAAGTTTTCGTTGTTAATGCTGATGATGCTTAAACAATGCCGATTCGCGTGGCGAT	67
Qy	229	ValProCysTyrSerPheGlyLeuAsnLysProAspPheLysAlaPheGlyLeuIleGlu	24
Db	676	GAACGCTGCGTACGTTGGCGTCAACATGGGTGACTATCAC-----CTGAAATCAT	72

D'b

Qy	249	GluspglyGlnLysTyrLeuAlaPheGlnPheaspLysLeuLeuProValGlyGluLeu	268
Db	727	CAGCAGGCGAAACCTGCTCGCGGTAAAGGGCAGAAAGTGTGAATGTGAAGAGATG	786
Qy	269	LysIleArgGlyAlaHisAsnTyrSerAsnAlaLeuAlaLeuAlaLeuGlyHisAla	288
Db	787	ANACTTTCCGGGACGATACACTACCAATACGCTGGCGCGCTGGCGTGGCAGATGCT	846
Qy	289	ValGlyLeuProPheAspAlaMetLeuGlyAlaLeuLysAlaPheSerGlyLeuAlaHis	308
Db	847	GCAGGGTTACGCGCTGCAGCAGCGCTGAAGCGTTAACCACTTCACTGCTGCCGCAT	906
Qy	309	ArgCysGlnTrpValArgGluArgGlnGlyValSerTyrTyrAspSerLysAlaThr	328
Db	907	CGCTTTCAAGTGTGCTGGAGCATACGCGGTACGTTGGATTACGATTTCGAAGCGACC	966
Qy	329	AsnValGlyAlaAlaLeuAlaAlaIleGluGlyLeuGlyAlaAspIleaspGlyLysLeu	348
Db	967	AACGTCGCGAGTACGGAAGCGGCGCTGAATGGCGTG-----CACGTAGACGCACATG	1020
Qy	349	ValLeuLeuAlaGlyGlyAspGlyLysGlyAlaAspPheHisAspLeuArgGluProVal	368
Db	1021	CATTGTGTGCTGGGTGCGATGTTAAATCGCGGACTTT-----AGCCCACTG	1068
Qy	369	AlaArgPheCysArgAla-----ValValLeuLeuGlyArgAspAlaGlyLeu	384
Db	1069	GCGGTACTCTGAATGCGGATACAGTACGTCTGTATTGTTTCGGTCTGACGCGCGCAG	1128
Qy	385	IleAlaGlnAlaLeuGlyAsnAlaValProLeuValArgValAlaThrLeuAspGluAla	404
Db	1129	CTGGCG---GCGCTACGCCCGGAAGTG---GCAGAACAAACCGAACTATGGAACAGCG	1182
Qy	405	ValArgGlnAlaAlaGluLeuAlaAArgGluGlyAspAlaValLeuLeuSerProAlaCys	424
Db	1183	ATGCGCTTGTGCTCGCGTGTTCAGCGCGGCGATGTTCTGCTCTCCCACTGCT	1242
Qy	425	AlaSerLeuAspMetPheLysAsnPheGluGluArgGlyArgLeuPheAlaLysAlaVal	444
Db	1243	GCCAGCCTTCAGTTCAGAACTTTGAACACACGAGGCAATCAGTTTGCCTCTGGCG	1302
Qy	445	GluGluLeu 447	
Db	1303	AAAGAGTTA 1311	
RESULT 5			
AAS52295			
ID	AAS52295	standard; DNA; 1317 BP.	
XX			
AC	AAS52295;		
XX			
DT	13-FEB-2002	(first entry)	
XX			
DE		E. coli DNA for cellular proliferation protein #17.	
XX			
KW		Antisense; ds; prokaryotic cellular proliferation gene;	
KW		antibiotic; antibacterial; drug design.	
XX			
OS		Escherichia coli.	
XX			
PN	WO200170955-A2.		
XX			
PD	27-SEP-2001.		
XX			
PF	21-MAR-2001; 2001WO-US09180.		
XX			
PR	21-MAR-2000; 2000US-191078P.		
PR	23-MAY-2000; 2000US-206848P.		
PR	26-MAY-2000; 2000US-20727P.		
PR	23-OCT-2000; 2000US-242578P.		
PR	27-NOV-2000; 2000US-253625P.		
PR	22-DEC-2000; 2000US-257931P.		
PR	16-FEB-2001; 2001US-269308P.		
XX			

(ELIT-) ELITRA PHARM INC.

Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;  
Yamamoto RT, Xu HH;  
WPI: 2001-611495/70.  
P-PSDB; AAU34436.

New polynucleotides for the identification and development of  
antibiotics, comprise sequences of antisense nucleic acids -

Claim 27; Seq ID No 5932; 51lpp; English.

The invention relates to antisense inhibitors of genes essential to  
prokaryotic cellular proliferation, their use in identifying the  
genes, their use in the discovery of novel antibiotics, the essential  
genes themselves and the encoded proteins. The prokaryotes used are  
Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella  
pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The  
invention is also useful for the identification of potential new targets  
for antibiotic development. The antisense nucleic acids can also be used  
to identify proteins used in proliferation, to express these proteins,  
and to obtain antibodies capable of binding to the expressed proteins.  
The proteins can be used to screen compounds in rational drug discovery  
programmes. The antisense nucleic acid sequence is also useful to screen  
for homologous nucleic acids which are required for cell proliferation in  
a wide variety of organisms. The present sequence encodes an  
essential prokaryotic cellular proliferation protein.

Note: The sequence data for this patent did not form part  
of the printed specification, but was obtained in electronic  
format directly from WIPO at  
ftp.wipo.int/pub/published\_pct\_sequences.

Sequence 1317 BP; 294 A; 323 C; 400 G; 300 T; 0 other;

Alignment Scores:

Pred. No.:	2.98e-84	Length:	1317
Score:	985.50	Matches:	227
Percent Similarity:	64.33%	Conservative:	58
Best Local Similarity:	51.24%	Mismatches:	137
Query Match:	43.61%	Indels:	21
DB:	23	Gaps:	9

US-09-701-229-2 (1-448) x AAS52295 (1-1317)

Qy	11	IleValValGlyLeuGlyLysSerGlyMetSerLeuValArgTyrLeuAlaArgArgGly	30
Db	28	GTCATTATCGCGCTGGCGCTCACCGGGCTTCTCGGTGGAGCTTTTCTCCTCGCTCGCGGT	87
Qy	31	LeuProPheAlaValAlaSerThrArgGluAsnProProGluLeuAlaThrLeuArgAla	50
Db	88	GTGACGGCGCGCGTATGGATACGGGTATGACACCGCTGGCTGGATAAATTA	141
Qy	51	GlnTyrProGln---ValGluValArgCysGlyGluLeuAspAlaGluPheLeuCysSer	69
Db	142	-----CCCGAAGCGGTAGACGCCACACGGGCACTCTGAATGATGATGGCTGATGGCG	195
Qy	70	AlaArgGluLeuTyrValSerProGlyLeuSerLeuArgThrProAlaLeuValGlnAla	89
Db	196	CGACATCTGATGTGCGGCAGTCCCGGTATTGACATGGCGCATCCATCTTAAGCGCTGCC	255
Qy	90	AlaAlaLysGlyValArgIleSerGlyAspIleAspLeuPheAlaArgGluAlaLysAla	109
Db	256	GCTGATGCGGAATCGAAATCGTTGGCGCATATCGAGCTGTCTGTGCGGAAGCACAAAGCA	315
Qy	110	ProIleValAlaIleThrGlySerAsnAlaLysSerThrValThrThrLeuValGlyGlu	129
Db	316	CCGATTGTGGCGATTACCGGTCTCAACGCCAAACACCGGTACACACGCTAGTGGGTGAA	375
Qy	130	MetAlaValAlaAlaAspLysArgValAlaValGlyGlyAsnLeuGlyThrProAlaLeu	149
Db	376	ATGCGAAGACGGCGGGGTAAAGCTGGTGGGTGGCAATATATGGCTGCTGCTGGTGGT	435



150	AspLeuLeuAlaAspAlleGluLeuTyrValLeuGluLeuSerPheGlnLeuGlu	169
QY		
436	ATGCTTACTGGATGATGAGTGTGAACGTGTACGTCTGGAACATGTCGAGCTGCCAGCTGGAA	495
DB		
170	ThrCysAspArgLeuAsnAlaGluValAlaThrValLeuAsnValSerGluAspHisMet	189
QY		
496	ACCACCTCCAGCTTACAGCGGTAGCAGCGCCATCTGAACGTGACTGAAGATCATATG	555
DB		
190	AspArgTyrAsp---GlyMetAlaAspTyrHisLeuAlaLysHisArgIlePheArgGly	208
QY		
556	GATCGCTATCCGTTTGGTTTACAACAGTATCGTCGACGAAACATGCGCATTTACGAAAC	615
DB		
209	AlaArgGlnValValValAsnArgAlaAspAlaLeuThrArgProLeuIleAlaAspThr	228
QY		
616	GCAGAAAGTTTGCCTGTTTAAATGCTGATGATGCCCTTAACATGCGCATTCGCGTGGCGAT	675
DB		
229	ValProCysTrpSerPheGlyLeuAsnLysProAspPheLysAlaPheGlyLeuIleGlu	248
QY		
676	GAACGCTGCGTACGCTTGGCGTCAACATGGGTGACTATCAC-----CTGAATCAT	726
DB		
249	GluAspGlyGlnLysTrpLeuAlaPheGlnPheAspLysLeuLeuProValGlyGluLeu	268
QY		
727	CAGCAGGGCGAAACCTGGCTGCGGGTTAAAGGGCAGAAAGTGTGTAATGTGAAGAGATG	786
DB		
269	LysIleArgGlyAlaHisAsnTyrSerAsnAlaLeuAlaLeuAlaLeuGlyHisAla	288
QY		
787	AAACTTCCGGGAGCATAACTACACCAATGCGCTGCGCGCTGCGCTGGCAGATGCT	846
DB		
289	ValGlyLeuProPheAspAlaMetLeuGlyAlaLeuLysAlaPheSerGlyLeuAlaHis	308
QY		
847	GCAGGGTTACCGGTCGCAGCACCTCGAAGCGCTTAACNCATTCATGCTGTCGCGCAT	906
DB		
309	ArgCysGlnTrpValArgGluArgGlnGlyValSerTyrTrpAspAspSerLysAlaThr	328
QY		
907	CGCTTTGAAGTCTGCTGGAGCATAAACGGCGTACGTTGGATTAAACGATTTCGAAAGCGACC	966
DB		
329	AsnValGlyAlaAlaLeuAlaAlaIleGluGlyLeuGlyAlaAspIleAspGlyLysLeu	348
QY		
967	AACGTCGGCAGTACGGAAGCGCGCTGAATGGCGCTG-----CACGTAGACGCGACATG	1020
DB		
349	ValLeuLeuAlaGlyGlyAspGlyLysGlyAlaAspPheHisAspLeuArgGluProVal	368
QY		
1021	CATTGTGCTGGTGGCGATGTGAATCGCGCGACTTT-----ACGCCACTG	1068
DB		
369	AlaArgPheCysArgAla-----ValValLeuLeuGlyArgAspAlaGlyLeu	384
QY		
1069	GCSCGTTACCTGAATGCGGATACGTAACGTAACGTCCTATTGTTTCGTCGACGCGCGCAG	1128
DB		
385	IleAlaGlnAlaLeuGlyAsnAlaValProLeuValArgValAlaThrLeuAspGluAla	404
QY		
1129	CTGGCG---CGCTACGCGCGGAAGTG---GCAGAACAAACCCGAAATATGGAACAGCGC	1182
DB		
405	ValArgGlnAlaAlaGluLeuAlaArgGluGlyAspAlaValLeuLeuSerProAlaCys	424
QY		
1183	ATCGCGTTGCTGCTCCGCGTGTTCACCGCGGCGATGTTGCTCTCTCCCCAGCGCTGT	1242
DB		
425	AlaSerLeuAspMetPheLysAsnPheGluGluArgLeuPheAlaLysAlaVal	444
QY		
1243	GCCAGCGCTTCATCAGTTCAGAACCTTGAACACACGAGGCAATGAGTTTGGCGGCTCGCGG	1302
DB		
445	GluGluLeu	447
QY		
1303	AAGGAGTGA	1311
DB		

RESULT 6

AAF28548/c

ID AAF28548 standard; DNA; 96109 BP.

XX

AAF28548;

XX

AC

XX

DT

04-APR-2001 (first entry)

XX

Genomic fragment #35.

```

XX Genomic library; bacteria; human upper airway; otitis media; sinusitis;
KW bronchopulmonary; endocarditis; meningitis; ss.
XX
XX Moraxella catarrhalis.
XX
XX WO200078968-A2.
XX
XX 28-DEC-2000.
XX
XX 16-JUN-2000; 2000WO-US16649.
XX
XX 18-JUN-1999; 99US-0140121.
XX
XX (INCY-) INCYTE GENOMICS INC.
XX
XX Lagace RE, Patterson C, Berg KL;
XX
XX WPI; 2001-041427/05.
XX
XX Genomic library for identifying diagnostic and therapeutic
PT compositions, and for identifying virulence factors, regulatory
PT elements and drug targets, comprises Moraxella catarrhalis nucleic
PT acids.
XX
XX Claim 1; Page 345-368; 545pp; English.
XX
XX The present invention relates to a Moraxella catarrhalis genomic library
CC comprising of a combination of 41 nucleic acid molecules (see
CC AAF28514-AAF28554). The library has a number of uses described in the
CC specification e.g. is useful for identifying diagnostic and therapeutic
CC compositions. M. catarrhalis (Branhamella catarrhalis) is a large
CC aerobic, gram-negative diplococcus, normally found among the bacterial
CC flora of human upper airways. M. catarrhalis is known to cause acute,
CC localised infections such as otitis media, sinusitis and bronchopulmonary
CC infection and life-threatening, systemic diseases including endocarditis
CC and meningitis.
XX
XX SQ Sequence 96109 BP; 28783 A; 18910 C; 20341 G; 28075 T; 0 other;

Alignment Scores:
Pred. No.: 1 17e-72 Length: 96109
Score: 889.50 Matches: 213
Percent Similarity: 60.43% Conservative: 68
Best Local Similarity: 45.81% Mismatches: 147
Query Match: 39.36% Indels: 37
DB: 22 Gaps: 12

US-09-701-229-2 (1-448) x AAF28548 (1-96109)
QY 12 ValValGlyLeuGlyTyrSerGlyMetSerLeuValArgTyrLeuAlaArgArgGlyLeu 31
Db 12721 GTCGTGGTGTGGGACGCTCGACGTGTGAATTTTGGTCAATCAGGTCAC 12662
QY 32 PropheAlaValAlaValAspThrArgGluAsnProProGluLeuAlaThrLeuAlaGln 51
Db 13661 AAAGTTAGCATCATCGATGTAATCCCAATCCT-----ACCTGGCAGATAAA 12614
QY 52 TyrProGln---ValGluValArgCysGlyGluLeuAlaGluPheLeuCysSerAla 70
Db 12613 CTCGCCACAGCGGCTTCATTGTCTATTTGGCAGCTTGGCAGATGAACCTTTAAAGCTACGCC 12554
QY 71 ArgGluLeuTyrValSerProGlyLeuSerLeuArgThrProAlaLeuValGlnAlaLa 90
Db 12553 GATCAGATCATCATACCCCGGCANTTAACCCAGAACGCCGAGTATGTGCGCCGCCAA 12494
QY 91 AlaLysGlyValArgIleSerGlyAspIleAspLeuPheAlaArgGluAlaLysAla--- 109
Db 12493 GCTCAAGGCATTCTCTGTGATCAGCGATGTTCAATTGTTTCATTGATACACTCAAGACGCA 12434
QY 110 -----ProIleValAlaIleThrGlySerAsnAlaLysSer 121
Db 12433 GATCTAACTCAAGCAACAACACACCAATCATCTCGCATTCACAGGCTCAATGCCAAAGC 12374

```



Qy	12	valValglyLeuGlyLysSerGlyMetSerLeuValArgTyrLeuAlaArgArgGlyLeu	31
Db	31	ATCATCGGGTGGCAAAACAGGTCTTCTTGTTGGATTATCTCTTATCCCAACAGGCT	90
Qy	32	ProPheAlaValValAspThrArgGluAsnProProGluLeuAlaThrLeuArgAlaGln	51
Db	91	AAATTCGTGTGATGTACTCGAAAATAATCCTACTGGTATTGATAAATTT-----	141
Qy	52	TyrProGln---ValGluValArgCysGlyGluLeuAspAlaGluPheLeuCysSerAla	70
Db	142	--CCTCAAAATATCCCTCTTCATACGTGGTAGTTAAATCAGGAATGGTTACTTGAAGC	198
Qy	71	ArgGluLeuTyrValSerProGlyLeuSerLeuLeuArgThrProAlaLeuValGlnAlaLa	90
Db	199	GATATGATTGTATTACCCAGGGCTTGGGTGCGGTAAAAACACAGCAAAATTCAAACGCGACTT	258
Qy	91	AlaLysGlyValArgIleSerGlyAspIleAspLeuPheAlaArgGluAlaLysAlaPro	110
Db	259	AAACGGGAGTGGNAGTAATCGCGCATTTGAATATTCTGCCGCGCAGCGACAAAGCCA	318
Qy	111	IleValAlaIleThrGlySerAsnAlaLysSerThrValThrThrLeuValGlyGluMet	130
Db	319	ATTCTGGGATTACAGGTTCAAAATGGTAAAGTACCGTAACTACTCTTTAGTTTATGAAATG	378
Qy	131	AlaValAlaAlaAspLysArgValAlaValAlaGlyAsnLeuGlyThrProAlaLeuAsp	150
Db	379	GCGAAAGCTGCTGGTGTGAAGTTGGTATGGCGGAAATATGGGATTCGCGTCTGTCTCA	438
Qy	151	LeuLeuAlaAspIleGluLeuTyrValLeuGluLeuSerSerPheGlnLeuGluThr	170
Db	439	TTGTTGAATGAAGATTGTGAACCTTAATGCTAGAGCTTCTAGTTTACGCTTGAGACA	498
Qy	171	CysAspArgLeuAsnAlaGluValAlaThrValLeuAsnValSerGluAspHisMetAsp	190
Db	499	ACTTATAGCTTAAAGCTGCGGACGACTGTCTTGAACGTGACTGAAGATCATATGGAT	558
Qy	191	ArgTyrAspGlyMetAlaAspTyrHisLeuAlaLysHisArgIlePheArgGlyAlaArg	210
Db	559	CGCTATATGATTTAGAAGATTATCGCCACAGCAAAATTCAGCATTTATCATATGCTTAAA	618
Qy	211	GlnValValValAsnArgAlaAspAlaLeuThr-----ArgProLeuIleAla	226
Db	619	GTAGGTGTGTGAACANTGAAGTAGCTGACTTTGGGGAACGAAAATCAAGCGAAA	678
Qy	227	AspThrValProCysTyrSerPheGlyLeuAsnLysProAspPheLysAlaPheGlyLeu	246
Db	679	CATACCGCTT-----TCTTTTGGCGAAAATAGTCGGAT-----TATTGGCTA	720
Qy	247	IleGluLysGlyGlnLysTyrLeuAlaPheGlnPheAspLysLeuLeuProValGly	266
Db	721	AAACTGAAATGGCAAGCAATATTAAATGGTAAAGATGAAGTGATTTTACCTTGTGAA	780
Qy	267	GlyLeuLysIleArgGlyAlaHisAsnTyrSerAsnAlaLeuAlaLeuAlaLeuGly	286
Db	781	GAAGCTACATTGGTTGGTCCCAATAATTATGAACATTTTGGCAGCAGCAGCATTGCCA	840
Qy	287	HisAlaValGlyLeuProPheAspAlaMetLeuGlyAlaLeuLysAlaPheSerGlyLeu	306
Db	841	CAAGCTATAGGTATTAATTTAGATTTCAATTCGTACCGCATTCGTCTATTTCAAAGGGTTA	900
Qy	307	AlaHisArgCysGlnTyrPValArgGluArgGlnGlyValSerTyrTyrAspAspSerLys	326
Db	901	GATCATCGTTTTCANTTAGTGCATCAGCTAATGGCAATCGTTGGATTAATGACTCTAAA	960
Qy	327	AlaThrAsnValGlyAlaAlaLeuAlaIleGluGlyLeuGlyAlaAspIleAspGly	346
Db	961	GCAACAAATGTGGGAGTACAGTTGCTGCATTTGGCTGGCTT-----TATATTGAGGCT	1014
Qy	347	LysLeuValLeuLeuAlaGlyLysGlyLysGlyAlaAspPheHisAspLeuArgGlu	366
Db	1015	AAATPGCATTTGTTGCTTAGCGGACGAGGAAAGGGCTGATTTTTCAGAAATTAGCTGAA	1074

Qy	367	ProValalaArg-----PheCysArgAlaValValLeuLeuGlyArgAsp	381
	:::	:::	
Db	1075	TTAATTATCAACACACACATTATTGTATTGT	1119
		:::	
		TTTGGTCGAGAT	
Qy	382	AlaGlyLeuIleAlaGlnAlaLeuGlyAsnAlaValProLeuValArgValAlaThrLeu	401
Db	1120	GGTCGGCTGCTGCAGAAATTTTCATCCGAAGTTAT	1158
		:::	---
Qy	402	AspGluAlaValArgGlnAlaAlaGluLeuAlaArg-----GluGlyAspAla	417
	:::		
Db	1159	TTGCATACAAATGGAACAAGCAGATAGATTTTTACGCCCAACATTGCAACGGGAGATG	1218
Qy	418	ValLeuLeuSerProAlaCysAlaSerLeuAspMetPheLysAsnPheGluGluArgGly	437
Db	1219	GTATTATTGTCGCTGCTGTGTGCAAGTCTCGATCAGTTTGCTTCTTTTGAAGCGCGC	1278
Qy	438	ArgLeuPhe	440
Db	1279	GAAGAAATTT	1287
RESULT 8			
AAAT42063			
ID	AAAT42063	standard; DNA; 1830121 BP.	
XX	AAAT42063;		
XX		14-SEP-1999 (first entry)	
XX		Haemophilus influenzae complete genome sequence.	
DE		Genome; bacterium; Haemophilus influenzae; computer readable medium;	
XX		expression modulating fragment; regulation; gene expression; vector;	
KW		organism; open reading frame; ORF; ds.	
XX		Haemophilus influenzae.	
OS			
XX			
PN	WO9633276-A1.		
XX		24-OCT-1996.	
PD			
XX		22-APR-1996; 96WO-US05320.	
PF			
XX		07-JUN-1995; 95US-0487429.	
PR		21-APR-1995; 95US-0426787.	
PR		07-JUN-1995; 95US-0476102.	
XX		(HUMA-) HUMAN GENOME SCI INC.	
PA		(UJJO ) UNIV JOHNS HOPKINS.	
XX			
PI	Adams MD, Fleischmann RD, Smith HO, Venter JC, White O;		
XX			
DR	WPI; 1996-485782/48.		
XX			
PT	Haemophilus influenzae Rd genome recorded on computer readable		
PT	medium - useful for identifying commercially important nucleic acid		
PT	fragments by homology searching		
XX			
PS	Claim 1; Page 77.2-77.1091; 1291pp; English.		
XX			
CC	This sequence represents the complete genome sequence of the bacterium		
CC	Haemophilus influenzae strain Rd. The invention relates to a computer		
CC	readable medium (CRM) having recorded upon it the complete H.influenzae		
CC	nucleotide sequence (I), a representative fragment of (I) or a nucleotide		
CC	sequence at least 99% identical to (I). By providing the full-length		
CC	genomic sequence in a computer readable form, it is possible to identify		
CC	commercially important nucleic acid fragments and expression modulating		
CC	fragments (EMFs) of the Haemophilus genome. The EMFs can be used to		
CC	regulate the expression of a nucleic acid molecule. Vectors and altered		
CC	organisms comprising the predicted ORFs can be used to produce any of the		
CC	polypeptide fragments of the H. influenzae Rd genome.		
XX	Sequence 1830121 BP; 567399 A; 350615 C; 347389 G; 564036 T; 682 other;		

**Alignment Scores:**

Pred. No.:	5.64e-70	Length:	1830121.
Score:	879.00	Matches:	206
Percent Similarity:	60.72%	Conservative:	63
Best Local Similarity:	46.50%	Mismatches:	136
Query Match:	38.89%	Indels:	38
DB:	17	Gaps:	10

US-09-701-229-2 (1-448) x AAT42063 (1-1830121)

Qy	12	ValValGlyLeuGlyLysSerGlyMetSerLeuValArgTyrLeuAlaArgGlyLeu	31
Db	1203162	ATCATCGGGCTTGGCAAAACAGGCTCTTCTGTGTGATATATCTCTATCCCAACAGGCT	1203221
Qy	32	ProPheAlaValValAspThrArgGluAsnProGluLeuAlaThrLeuArgAlaGln	51
Db	1203222	AATATTCGTGTGATGTGATATCGAAATAATCCCTACTGGTATTGATAAAGCTT-----	1203272
Qy	52	TyrProGln---ValGluValAlaGlyGluLeuAspAlaGluPheLeuCysSerAla	70
Db	1203273	---CCTCAAAATATCCCTCTTCATCATCTGGTAGTTAAATCATCAGGAATGGTTACTTGAAGC	1203329
Qy	71	ArgGluLeuTyrValSerProGlyLeuSerLeuArgThrProAlaLeuValGlnAlaLa	90
Db	1203330	GATATGATGTTTATAGCCAGGCGCTTGGCGTAAAAACACACAGAAATTCAAACCGCACTT	1203389
Qy	91	AlaLysGlyValArgTleSerGlyAspIleAsnLeuPheAlaArgGluAlaLysAlaPro	110
Db	1203390	AAACGGGAGTGAAGTATCGCGATTTGAATTTATTCGCCGCCAGCGCAAAAGCCA	1203449
Qy	111	IleValAlaThrGlySerAsnAlaLysSerThrValThrThrLeuValGlyGluMet	130
Db	1203450	ATTGTGGGATTCACAGGTTCAATGGTTAAAGTACCGTAACCTACTTACTTTATGAATG	1203509
Qy	131	AlaValAlaAlaAspLysArgValAlaValAlaGlyLysLeuGlyThrProAlaLeuAsp	150
Db	1203510	GCAAAAGCTCTCGTGTGAAGTGTGTGGCGGAAATATTTGGGATTCOCGCTTTGTCA	1203569
Qy	151	LeuLeuAlaAspAspIleGluLeuTyrValLeuGluLeuSerPheGlnLeuGluThr	170
Db	1203570	TTGTTGAATGAAGATTGTGAACCTTTATGTACTAGAGCTTTCAGTTTTCAGCTTGAGACA	1203629
Qy	171	CysAspArgLeuAsnAlaGluValAlaThrValLeuAsnValSerGluAspHisMetAsp	190
Db	1203630	ACTTATAGCTTAAAGCTGGCGCAGCAGCTGTCTTGAACGTGACTCAAGATCATATGGAT	1203689
Qy	191	ArgTyrAspGlyMetAlaAspTyrHisLeuAlaLysHisargTlePheargGlyAlaArg	210
Db	1203690	CGCTATATGATTATAGAAATATTCGCCAAGCAAAATTTACGCTATTTATCATATGCTAA	1203749
Qy	211	GlnValValValAsnArgAlaAspAlaLeuThr-----ArgProLeuIleAla	226
Db	1203750	GTAGGTGTGTTGAACATGAAGATAGCGTGACTTTTGGGGGAAACGAAATCAGCGCAA	1203809
Qy	227	AspThrValProCysTrpSerPheGlyLeuAsnLysProAspPheLysAlaPheGlyLeu	246
Db	1203810	CATACCGTT-----TCCTTTGGCGAAATAGTCCGAT-----TATTGGCTTA	1203851
Qy	247	IleGluLusAspGlyGlnLysTrpLeuAlaPheGlnPheAspLysLeuLeuProValGly	266
Db	1203852	AAAACCTGAAATGGCAAGCAATATTTAAATGGTAAAAAGATGAAGTGATTTTACCTTGTGAA	1203911
Qy	267	GluLeuLysIleArgGlyAlaHisAsnTyrSerAsnAlaLeuAlaLeuAlaLeuGly	286
Db	1203912	GAAGCTACATTTGTTGGCTGCCAATATATATGAACATTTTGGCACACACACATTTGGCA	1203971
Qy	287	HisAlaValGlyLeuProPheAspAlaMetLeuGlyAlaLeuLysAlaPheSerGlyLeu	306
Db	1203972	CAAGCTATAGGTATTAATTAGATTCAATTCGTACCGCACTTCGTCATTTCAAGGGTTA	1204031
Qy	307	AlaHisArgCysGlnTrpValArgGluArgGlnGlyValSerTyrTrpAspSerLys	326
Db	1204032	GATCATCGTTTCAATTAGTCATCAAGCTAATGATGATTCGTTGGATTAATGACTTAA	1204091

Qy	327	AlaThrAsnValGlyAlaAlaLeuAlaAlaIleGluGlyLeuGlyAlaAspIleAspGly	346
Db	1204092	GCACAAATGTGGGAGTACATGCTGCTGCTT-----TATATTGAGGCT	1204145
Qy	347	LysLeuValLeuLeuAlaGlyGlyAspGlyLysGlyAlaAspPheHisAspLeuArgGlu	366
Db	1204146	AAATTCATTGTGTCTAGCGGAGACGGAAAGGGCTGATTTTCAGAAATTAGCTGAA	1204205
Qy	367	ProValAlaArg-----PheCysArgAlaValValLeuLeuGlyArgAsp	381
Db	1204206	TTRATTAAATCAACACACATATTGTTGTTATGTT-----TTTGTCGAGAT	1204250
Qy	382	AlaGlyLeuIleAlaGlnAlaLeuGlyAsnAlaValProLeuValArgValaIaThrLeu	401
Db	1204251	GGTCCGCTGCTGCATAAATTTTCATCCGCAAGTTAT-----TTG	1204289
Qy	402	AspGluAlaValArgGlnAlaAlaGluLeuAlaArg-----GluGlyAspAla	417
Db	1204290	TTGCATACATGGACACAGCGATAGATTTTACGCCCAACATGCNAACGGAGATATG	1204349
Qy	418	ValLeuLeuSerProAlaCysAlaSerLeuAspMetPheLysAsnPheGluGluArgGly	437
Db	1204350	GTATTATGTGCGCTGCTGTCGAAGTCTCGATCAGTTTCTGCTCTTTTGAAGACGCGC	1204409
Qy	438	ArgLeuPhe 440	
Db	1204410	GAAGAAATTT 1204418	
RESULT	9		
ID	ABA92787/c		
XX	ID	ABA92787 standard; DNA; 640681 BP.	
AC	ABA92787;		
XX	27-MAR-2002 (first entry)		
XX	Buchnera sp. genomic DNA SEQ ID NO:1.		
XX	Buchnera; cockroach-symbiotic bacterium; cockroach extermination;		
KW	circular; ds.		
XX	Buchnera sp.		
OS	JP2001292771-A.		
PN	23-OCT-2001.		
XX	07-APR-2000; 2000JP-0107160.		
XX	07-APR-2000; 2000JP-0107160.		
XX	(RIKA ) RIKAGAKU KENKYUSHO.		
PA	WPI; 2002-126043/17.		
XX	A genomic DNA of cockroach-symbiotic bacterium		
PT	claim 1; Page 16-230; 237pp; Japanese.		
XX	The present invention describes a gene (I) derived from Buchnera sp.		
CC	containing the DNA (a) or (b), (a) has a fully defined base pair		
CC	sequence selected from a table of sequences found in the Buchnera sp.		
CC	genomic DNA of ABA92787 given in the specification or is a DNA selected		
CC	from complementary DNA sequences, and (b) is a DNA which hybridises with		
CC	the DNA (a) and encodes a protein. Also described are: (1) a recombinant		
CC	vector (iii) containing (1); (2) a transformant (iii) containing (ii);		
CC	(3) a genomic DNA of Buchnera sp. containing the sequence given in		
CC	ABA92787; (4) a plasmid derived from Buchnera sp. containing DNA (c) or		
CC	(d), (c) is a DNA containing a fully defined sequence given in ABA92788		
CC	or ABB92789 and (d) is a plasmid which hybridises with a DNA; and (5) a		
CC	method for the preparation of a protein in which (iii) is cultured and		
CC	the expression protein of the objective protein is collected from the		

# A genomic DNA of cockroach-symbiotic bacterium

claim 1: page 16-230: 237pp: Japanese.

xx The present invention describes a gene (I) derived from *Buchnera* sp. containing the DNA (a) or (b). (a) has a fully defined base pair sequence selected from a table of sequences found in the *Buchnera* sp. genomic DNA of ABA92787 given in the specification or is a DNA selected from complementary DNA sequences, and (b) is a DNA which hybridises to the DNA (a) and encodes a protein. Also described are: (1) a recombinant vector (II) containing (I); (2) a transformant (III) containing (II); (3) a genomic DNA of *Buchnera* sp. containing the sequence given in ABA92787; (4) a plasmid derived from *Buchnera* sp. containing DNA (c) or (d), (c) is a DNA containing a fully defined sequence given in ABA92789 and (d) is a plasmid which hybridises with a DNA; and (5) a method for the preparation of a protein in which (III) is cultured and the expression protein of the objective protein is collected from the

CC resultant culture. The DNA is useful for developing agricultural  
 CC chemicals for exterminating cockroaches. The present sequence represents  
 CC the specifically claimed Buchnera sp. genomic DNA sequence, from the  
 CC present invention.

XX  
 SQ Sequence 640681 BP; 237522 A; 83822 C; 84757 G; 234580 T; 0 other;

#### Alignment Scores:

Pred. No.: 4.74e-55 Length: 640681  
 Score: 716.00 Matches: 151  
 Percent Similarity: 57.50% Conservative: 102  
 Best Local Similarity: 34.32% Mismatches: 173  
 Query Match: 31.68% Indels: 14  
 DB: 24 Gaps: 7

US-09-701-229-2 (1-448) x ABA92787 (1-640681)

Qy 11 IleValValGlyLeuGlyLysSerGlyMetSerLeuValArgTyrLeuAlaArgArgGly 30  
 Db 241263 TTAATTTTAGGAATGGGTTTAAACAGGAATATCTTGCATTATATTTTAAAAAAGGA 241204  
 Qy 31 LeuProPheAlaValAlaAspThrArgGluAsnProGluLeuAlaThrLeuArgAla 50  
 Db 241203 ATAAACCTAAATATATGATGATCTAAACATCTTCAATTTTATTAATTT 241150  
 Qy 51 GlnTyrProGln---ValGluValArgCysGlyGluLeuAspAlaGluPheLeuCysSer 69  
 Db 241149 -----CCTCAAAATATGATATAGTTAGGAAGTTTAGACCATCAGTGATCTAGAA 241096  
 Qy 70 AlaArgGluLeuTyrValSerProGlyLeuSerLeuArgThrProAlaLeuValGlnAla 89  
 Db 241095 TCAGACTTAATTTGTTAAGTCCGGTATTTTCATTTAAACCTATTTTAAATTAAGCA 241036  
 Qy 90 AlaAlaLysGlyValArgIleSerGlyAspIleAspLeuPheAlaArgGluAlaLysAla 109  
 Db 241035 CGTTGTAGTATGATATATAGTATGATATGAACTATTTCTAGAGAGTACCTGT 240976  
 Qy 110 ProIleValAlaIleThrGlySerAsnAlaLysSerThrValThrThrLeuValGlyGlu 129  
 Db 240975 CCTATTATTTCAATAACAGGAGCTAATGGAAGAGTACTGTAGCTACCATGATTTGAAAAA 240916  
 Qy 130 MetAlaValAlaAlaAspLysArgValAlaValGlyGlyAsnLeuGlyThrProAlaLeu 149  
 Db 240915 ATTGCAAAAATCAGGATATAAGCTTTCTAGTGGTATATAGTGTTCCTCGTACTA 240856  
 Qy 150 AspLeuLeuAlaAspAspIleGluLeuTyrValLeuGluLeuSerSerPheGlnLeuGlu 169  
 Db 240855 GAAATACTTGATAAGAGCAGCATTTATACATATAGAACTGTCTAGTTTCAACTAGAA 240796  
 Qy 170 ThrCysAspArgLeuAsnAlaGluValAlaThrValLeuAsnValSerGluAspHisMet 189  
 Db 240795 AATACATTTAATTTAAATCTAAGATAGCAGTTATTTCTTAATATAGGAAGATCATATC 240736  
 Qy 190 AspArgTyr---AspGlyMetAlaAspTyrHisLeuAlaLysHisArgIlePheArgGly 208  
 Db 240735 AATCGATACCAATGGATGCCAACATATAAAACACTAAATGTCTGTGTTTAAATCAA 240676  
 Qy 209 AlaArgGlnValValAlaAsnArgAlaAspAlaLeuThrArgProLeuIleAlaAspThr 228  
 Db 240675 GCAGAAATTTCTATATTAATCTAATGATAAGATAGAAAAAGTCTTATTCATAGTAAG 240616  
 Qy 229 ValProCysTrp---SerPheGlyLeuAsnLysProAspPheLysAlaPheGlyLeuIle 247  
 Db 240615 AATAAAAAATGGATTAGTTTGGCAACAATAGAAAGTATTCGT-----ATTGT 240565  
 Qy 248 GluGluAspGlyGlnLysTrpLeuAlaPheGlnPheAspLysLeuLeuProValGlyGlu 267  
 Db 240564 TCTAAAGACGATGACCTATTTTATTTTAAAAATAAAAAATATTAATTAATCTAGTACAA 240505  
 Qy 268 LeuLysIleArgGlyAlaHisAsnTyrSerAsnAlaLeuAlaAlaLeuLysGlyHis 287  
 Db 240504 ATATTATATATGATATCATATTAATTAATAATATATATATAGTCGTCTTGTAGCTATTTTCAGAT 240445

Qy 288 AlaValGlyLeuProPheAspAlaMetLeuGlyAlaLeuLysAlaPheSerGlyLeuAla 307  
 Db 240444 GCAATGCAATCCCTAGAAATGACCAATAATGACTTAAAGATTTCTCGAATTTACCA 240385  
 Qy 308 HisArgCysGlnTyrPValArgGlnGlyValSerTyrTyrAspAspSerLysAla 327  
 Db 240384 CATCGATTTCAATAATAATAAAATCAAAAGAGGAGTACGTGGATAAAGGATCTAAATCC 240325  
 Qy 328 ThrAsnValGlyAlaAlaLeuAlaLalaLeuGluGlyLeuGlyAlaAspIleAspGlyLys 347  
 Db 240324 ACAATGTAAATAGTACTCAAGTGTCTTTAAATTTCTATA-----AAACAACGTGAACA 240271  
 Qy 348 LeuValLeuLeuAlaGlyGlyAspGlyLysGlyAlaAspPheHisAspLeuArgGluPro 367  
 Db 240270 ATAAGATTTATTATAGCGCGGATAGTAATCTGCAAAATTTTAAATATATATAAATAATTT 240211  
 Qy 368 ValAlaArgPheCysArgAlaValValLeuLeuGlyArgAspAlaGlyLeuIleAlaGln 387  
 Db 240210 TTTAGAACTTTTAAAAATAAAATTTACTGTGTTTGGAGAGATGGCATTTAAATTTGCAAAA 240151  
 Qy 388 AlaLeuGlyAsnAlaValProLeuValArgValAlaThrLeuAspGluAlaValArgGln 407  
 Db 240150 ATATGTGAAAAA-----AATCTATTTATGTTGAAAAATTTAAAAAAGCAGTATTTTA 240097  
 Qy 408 AlaAlaGluLeuAlaArgGluGlyAspAlaValLeuLeuSerProAlaCysAlaSerLeu 427  
 Db 240096 ATTTTCAAAACAACTCAATCAGTGTATACCTGTTCTTTGCTCTGCTGGTGCAGCAGCTTA 240037  
 Qy 428 AspMetPheLysAsnPheGluGluArgGlyArgLeuPheAlaLysAlaValGluLeuLeu 447  
 Db 240036 GGTCATTTTCTAATCTTTGAAGAGAGAGGTTAATCTTTTATATAAATTTAATAAGGAAATA 239977  
 RESULT 10  
 AAZ53175/c  
 ID AAZ53175 standard; DNA; 1380 BP.  
 XX  
 AC AAZ53175;  
 XX  
 DT 21-MAR-2000 (first entry)  
 XX  
 DE Neisseria meningitidis ORF 090 partial DNA sequence SEQ ID NO:301.  
 XX  
 KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;  
 KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;  
 KW antibacterial; gene therapy; ds.  
 OS Neisseria meningitidis.  
 XX  
 FN W09957280-A2.  
 XX  
 PD 11-NOV-1999.  
 XX  
 PF 30-APR-1999; 99WO-US09346.  
 XX  
 PR 01-MAY-1998; 98US-0083758.  
 PR 31-JUL-1998; 98US-0094869.  
 PR 02-SEP-1998; 98US-0098994.  
 PR 02-SEP-1998; 98US-0099062.  
 PR 09-OCT-1998; 98US-0103749.  
 PR 09-OCT-1998; 98US-0103794.  
 PR 09-OCT-1998; 98US-0103796.  
 PR 25-FEB-1999; 99US-0121528.  
 XX  
 XX (CHIR ) CHIRON CORP.  
 PA (GENO-) INST GENOMIC RES.  
 XX  
 XX Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;  
 PI Petersen J, Pizzo M, Rappuoli R, Ratti G, Scalato E, Scarselli M;  
 PI Tettelin H, Venter JC;  
 XX  
 WPI: 2000-062150/05.  
 DR P-PSDB; AAY74413.  
 XX

QY	11	IleValValGlyLeuGlyLysSerGlyMetSerLeuValArgTyrLeuAlaArgGly	30
Db	1341	CTGCTCGCGGACTTCGGCGGTACGGGTATTTCATGATTCCCTACCTGCGCAAAACGGC	1282
QY	31	LeuProPheAlaValValAspThrArgGluAsnProGluLeuAlaThrLeuArgAla	50
Db	1281	CGCGAGGTGCTCGGTATGATGGCGGAGCTGAACCCGGACGCGGTGCGCAAAATCGGTAAAG	1222
QY	51	GlnTyrProGlnValGluValArgCysGlyGluLeuAspAlaGluPheLeuCysSerAla	70
Db	1221	ATGTTTACGCGGTGTGTGTTTACAGGGCGCTCTGAAAGATCGCGTGGACAACGGTTTC	1162
QY	71	ArgGluLeuTyrValSerProGlyLeuSerLeuArgThrProAlaLeuValGlnAlaAla	90
Db	1161	GATATTCTGGCTCTAGTCCCGCATCAGCGCAGCGCGGATATTGAGCGGTTCAAG	1102
QY	91	AlaLysGlyValArgIleSerGlyAspIleAspLeuPheAla	105
Db	1101	CAAAACGGCGAGCGCGTGTGGCGACATCGAATTGCTGGCGGACATTTGTGAACCGCCGG	1042
QY	106	GluAlaLysAlaProIleValAlaIleThrGlySerAsnAlaLysSerThrValThrThr	125
Db	1041	GAGCACAAG----GTATTGCGATTACGGCGACGACGGCAAAACCGGTAACGAGC	988
QY	126	LeuValGlyGluMetAlaValAlaAlaAspLysArgValAlaValGlyGlyAsnLeuGly	145
Db	987	CTGGTCCGGTATCTCTGTATCAAGTCCGGCTGGATACGTTATCGCGGCAATATCGGC	928
QY	146	ThrProAlaLeuAspLeu-----LeuAlaAspAspIleGluLeuTyrValLeu	161
Db	927	ACGCGGTTTGGAGCGGAATTGGACCGCGAAGGCNAAGCGGACGCTGTGGGTGTG	868
QY	162	GluLeuSerSerPheGlnLeuGluThrCysAspArgLeuAsnAlaGluValAlaThrVal	181
Db	867	GAGCTTTTCCAGCTTCCAACTGAAACACCGAAAGCGCTCGGCGACGGGTG	808
QY	182	LeuAsnValSerGluAspHisMetAspArgTyrAspGlyMetAlaAspTyrHisLeuAla	201
Db	807	CTGAACATTTCCGAGAGCCATCTCGACCGCTACGACGACTTGCTCGCACTATCGGCATACC	748
QY	202	LysHisArgIlePheArgGlyAlaArgGlnValValAlaAsnArgAlaAlaLeuThr	221

PR 30-APR-1999; 99US-0132068.  
 XX (CHIR ) CHIRON CORP.  
 XX Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;  
 PI Masignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scariato V;  
 PI Rappuoli R, Pizza M;  
 XX WPI; 2000-318079/27.  
 XX Isolated nucleotide sequences of *Neisseria meningitidis* which can be  
 PT used in the diagnosis and treatment of *N. meningitidis* infection and  
 PT other *Neisseria* infections, for example, *N. gonorrhoea*.  
 XX Claim 7; Page 547-567; 1760pp; English.  
 XX The present invention describes methods of obtaining immunogenic  
 CC proteins from *Neisseria* genomic sequences. AAA81453 to AAA82414  
 CC represent specifically claimed *Neisseria meningitidis* genomic DNA  
 CC sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent  
 CC *Neisseria* DNA sequences and their corresponding proteins; AAA81254 to  
 CC AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the  
 CC isolation of *Neisseria meningitidis* DNA sequences; and AAA81322 to  
 CC AAA81452 represent *Neisseria meningitidis* MenB polynucleotide ORF  
 CC sequences, which are all used in the exemplification of the present  
 CC invention. The nucleic acid sequences, protein sequences, and antibodies  
 CC against them, can be used in the manufacture of a composition. The  
 CC composition can be used as a medicament (or in the manufacture of a  
 CC medicament) for treating, preventing or diagnosing infection due to  
 CC *Neisseria* bacteria. For example, some of the identified proteins could  
 CC be components of vaccines against *Meningococcus B*; against all serotypes;  
 CC and/or against all pathogenic *Neisseriae*. Identification of sequences  
 CC from the bacterium will also facilitate production of biological probes,  
 CC particularly organism-specific probes. Attempts to make efficacious  
 CC *Meningococcus B* vaccines have failed mainly due to antigen tolerance.  
 CC Multivalent vaccines have also been tried but none have successfully  
 CC overcome antigenic variability. The provision of further, complete  
 CC sequences may provide an opportunity to identify secreted or surface  
 CC exposed proteins that may be presumed targets for the immune system and  
 CC which are not antigenically variable or at least more conserved than  
 CC other more variable regions.  
 XX SQ Sequence 69936 BP; 16512 A; 19259 C; 17399 G; 16763 T; 3 other;  
 Alignment Scores:  
 Pred. No.: 3,58e-55 Length: 69936  
 Score: 704.00 Matches: 173  
 Percent Similarity: 53.52% Conservatives: 70  
 Best Local Similarity: 38.11% Mismatches: 177  
 Query Watch: 31.15% Indels: 34  
 DB: 21 Gaps: 8  
 US-09-701-229-2 (1-448) x AAA81479 (1-69936)  
 QY 11 IleValValGlyLeuGlyLysSerGlyMetSerLeuValArgTyrLeuAlaArgGly 30  
 DB 52618 CTGCTCCCGGACTCGCGGTATTCGATTTTCATGATTCCTGCGCAAAACGGC 52559  
 QY 31 LeuProPheAlaValAlaValSerThrArgGluAsnProProGluLeuAlaThrLeuArgAla 50  
 DB 52558 CGCGAGGTGCTCGGTATGATGCGGAGCTGAGCCGGAACCGCGTGCAGAAATCGGTAA 52499  
 QY 51 GlnTyrProGlnValGluValArgCysGlyGluLeuAspAlaGluPheLeuCysSerAla 70  
 DB 52498 ATGTTTACGGGTGGTGGTTTACACGGCGGCTGAAAGATCGCTGGACACGGTTTC 52439  
 QY 71 ArgGluLeuTyrValSerProGlyLeuSerLeuArgThrProAlaLeuGlnAlaAla 90  
 DB 52438 GATATTCGCTCTCAGTCCCGGCATCAGCGAGCGCGCGGATATTGAGCGTTTCAAG 52379  
 QY 91 AlaLysGlyValArgIleSerGlyAspIleAspLeuPheAla-----Arg 105  
 DB 52378 CAAAACGGCGAGCGGTGTTGGCGGACATCGAATTGCTGCGGACATTGTGAACCGCGG 52319

QY 106 GluAlaLysAlaProIleValAlaIleThrGlySerAsnAlaLysSerThrValThrThr 125  
 DB 52318 GAGCAGCAG-----GTAATTGCGATACCGCGCAGCAACGCAACCGGTAAACGAGC 52265  
 QY 126 LeuValcylGluMetAlaValAlaAlaAspLysArgValAlaValGlyGlyAsnLeuGly 145  
 DB 52264 CTGGTGGCTATCTCTGTATCAAGTGGCGGTGATACCTGTTATCGCGGCAATATCGGC 52205  
 QY 146 ThrProAlaLeuAspLeu-----LeuAlaAspAspIleGluLeuTyrValLeu 161  
 DB 52204 AC CGCGGTTTGGAGCGGGAATGCGACGCGGAGGCAAAAGCGGAGGTGGTGGTGTG 52145  
 QY 162 GluLeuSerSerPheGlnLeuGluThrCysAspArgLeuAsnAlaGluValAlaThrVal 181  
 DB 52144 GAGCTTTCAGCTTCCAACTGGAAACACCGCAAGCGCTGCGTCCGAGTGGCGGACGGTG 52085  
 QY 182 LeuAsnValSerGluAspHisMetAspArgTyrAspGlyMetAlaAspTyrHisLeuAla 201  
 DB 52084 CTGAACATTTCCGAAGACCATCTGACCGCTACGACGACTGTCTCGACTATGCGCATACC 52025  
 QY 202 LysHisArgIlePheArgGlyAlaArgGlnValValValAsnArgAlaAspAlaLeuThr 221  
 DB 52024 AAAGCCAGATTTTCCGTGGCGGCGGTGCGAGGTGTTGATGCGGAGCATGCGTTC--- 51968  
 QY 222 ArgProLeuIleAlaAspThrValProCysTyrPhePheGlyLeuAsnLysProAspPhe 241  
 DB 51967 -----TGCGCGCGATGAAGCGTGCAGCGTGGCGGCGGAGGTA 51935  
 QY 242 LysAlaPheGlyLeuIleGluAspGlyGlnLysTrpLeuAlaPheGlnPhe----- 259  
 DB 51934 AAATGGTTTTCGTTG---GAACACGAAGCTGATTCTGTGGACCGAGGAGCGCGC 51878  
 QY 260 -----AspLysLeuLeuProValGlyGluLeuLysIleArgGlyAlaHis 274  
 DB 51877 CTGAAACAGGCAATGAAGATTGTTGTCAGCGAACACATCCGTTGCAAGGTCTGCAC 51818  
 QY 275 AsnTyrSerAsnAlaLeuAlaLeuLeuGlyHisAlaValGlyLeuProPheAsp 294  
 DB 51817 AACCGCGTAACTGATGCTGCGTGTGTGTGAGGCGCATCGTTTCTCGCGGAA 51758  
 QY 295 AlaMetLeuGlyAlaLeuLysAlaPheSerGlyLeuAlaHisArgCysGlnTrpValArg 314  
 DB 51757 GCATGTCTCGAACACGCTCAAAACCTCCAGGCGTGGCGCGGCAAAATCGGC 51698  
 QY 315 GluArgGlnGlyValSerTyrTyrAspSerLysAlaThrAsnValGlyAlaAlaLeu 334  
 DB 51697 GAGAAACCGCGGTGTTGTTATCGACGACAGCAAGCGCATGTGCGCGCGCATGCC 51638  
 QY 335 AlaAlaIleGluGlyLeuGlyAlaAspIleAspGlyLysLeuValLeuAlaGlyGly 354  
 DB 51637 GCGCGGATTCGCGGTTG-----CAAAATCGCTCTCTCGTATTTGGCGGC 51590  
 QY 355 AspGlyLysGlyAlaAspPheHisAspLeuArgGluProValAlaAlaArgPheCysArgAla 374  
 DB 51589 ATGGGTAAAGCGAGGACTTCCACCGCTGCGCGCATGCGTGTAGGCAAGCGCAAGGCGC 51530  
 QY 375 ValValLeuLeuGlyArgAspAlaGlyLeuIleAlaGlnAlaLeu---GlyAsnAlaVal 393  
 DB 51529 GTGTTCTGATGTTGTCGATGCGCGCAATCCCGCGCATTTGGAGCGGTGGCGGTG 51470  
 QY 394 ProLeuValArgValAlaThrLeuAspGluAlaValArgGlnAlaGluLeuAlaArg 413  
 DB 51469 AATATGACGAGTCCGCCACTTTGGGAGAGCGCTTTCAGAGCGCATATGCTCAAGCGCAA 51410  
 QY 414 GluGlyAspAlaValLeuLeuSerProAlaCysAlaSerLeuAspMetPheLysAsnPhe 433  
 DB 51409 GCAGCGATATTGTTGCTCGCGCGCTGCGGAGCTTTGATATGTTTCAAGGCTAC 51350  
 QY 434 GluGluArgGlyArgLeuPheAlaLysAlaValGluLeuLeu 447  
 DB 51349 GCGCACCGTTCGGAAGTGTATTATCGAAGCGTTTAAAGCGTTTG 51308



RESULT 12

AAF21607  
ID AAF21607 standard; DNA; 349980 BP.

AA  
AC AAF21607;

DT 13-MAR-2001 (first entry)

DE Neisseria meningitidis B nucleotide sequence SEO ID NO:108.

Neisseria meningitidis; Neisseria gonorrhoeae; immunogenic; vaccine;  
diagnosis; antigen; detection; infection; gene therapy; antibacterial;  
ds.

XX  
05  
Neisseria meningitidis.XX  
PN WO200066791-A1XX  
PD  
09-NOV-2000.

08-MAR-2000: 2000WO-ITS05928-XX PF

XX 30-APR-1999: 99US-0132068.

PR 08-OCT-1999; 99WU-USZ3373;  
PR 28-FEB-2000; 2000GB-0004695;

PA (CHIR ) CHIRON CORP.

FA (GENO<sup>-</sup>) INSI GENOMIC RES:  
XX

PI Pizza M, Hickey E, Pete

PI Frazer CM, Grandt G;  
XX  
DR WPI; 2000-647603/62.

PT *Neisseria meningitidis*PS Claim 7; Appendix A; 692pp; English.  
XX  
XX frames are used to detect, create and prevent neisseria infections

CC The present invention describes the

to AAF21613 represent fragments of the NMB genomic sequences, as the sequence was too long to go in a record on its own it was split into 8 sequences which overlap each other at the beginning and end of each sequence by 49980 bp (i.e. the last 49980 bp of AAF21544 is repeated at the beginning of AAF21607, the last 49980 bp of AAF21547 is repeated at the beginning of AAF21608, and so on). AAF21545 to AAF21588 encode the *Neisseria* proteins given in AAB59550 to AAB59593, and AAF21589 to AAF21606 represent PCR primers which are used in the exemplification of the present invention. The NMB genome and fragments from it have antibacterial activity, and can be used in vaccines and gene therapy. *Neisseria* nucleic acids, proteins and/or antibodies which binds to the proteins can be used in compositions for treating or preventing infection due to *Neisseria* bacteria or as a diagnostic reagent for detecting the presence of *Neisseria* bacteria or of antibodies raised to *Neisseria* bacteria. Computers, computer memory, computer storage medium or computer databases can be used in a search to identify open reading frames (ORFs) or coding sequences within the NMB genome. The DNA sequences provide further opportunities to find antigenic or immunogenic proteins which are more effective in vaccines than the outer membrane proteins currently used.

Sequence 349980 BP: 84410 A: 84863 C: 94187 G: 86520 T: 0 other:  
XX

**Alignment Scores:**

Pred. No.:	2,976-54	Length:	349980
Score:	704.00	Matches:	173
Percent Similarity:	53.5%	Conservative:	70
Best Local Similarity:	38.1%	Mismatches:	177
Query Match:	31.1%	Indels:	34
DB:	21	Gaps:	8

US-09-701-229-2 (1-448) X AAF21607 (1-349980)

Qy	11	IleValValGlyLeuGlyLysSerGlyMetSerLeuValArgTyrLeuAlaArgArgGly	30
Db	128569	CTCGTCGCGGACTCGCGGGTACGGGTATTTCCATGATTCCTACGTCCGCAAAACGGC	128628
Qy	31	LeuProPheAlaValValAspThrArgGluAsnProGluLeuAlaThrLeuArgAla	50
Db	128629	CGCGAGGTGTGCTGGTATGATGCGGAGCTGAAGCCGAACGGCTGTGCGCAATCGGTAG	128688
Qy	51	GlnTyrProGlnValGluValArgCysGlyGluLeuAspAlaGluPheLeuCysSerAla	70
Db	128689	ATGTTTGACGGTGGTGTCTTACACGGCCGCTCTGAAGAGATGCGCTGCACACGGTTTC	128748
Qy	71	ArgGluLeuTyrValSerProGlyLeuSerLeuArgThrProAlaLeuValGlnAlaAla	90
Db	128749	GATATTCTGGCTTCAGTCCCGGATCAGCAGCGCAGCGGATATTGAGCGTTCAAG	128808
Qy	91	AlaLysGlyValArgIleSerGlyAspIleAspLeuPheAla	105
Db	128809	CAAAACGGGACCGGTGTGGCGCATCGAATTGCTGGCGGACATTTGTAACCGCCGG	128868
Qy	106	GluAlaLysAlaProIleValAlaIleThrGlySerAsnAlaLysSerThrValThrThr	125
Db	128869	GACGACAAG-----GTAATTGCGATTACCGCACAACGGCAAAACACCGTACGAGC	128922
Qy	126	LeuValGlyGluMetAlaValAlaAlaAspLysArgValAlaValGlyAsnLeuGly	145
Db	128923	CTGTGCGGTATCTCTGATCAAGTGGCGGTGGATACCGTTATCGGGGCANTATCGGC	128982
Qy	146	ThrProAlaLeuAspLeu-----LeuAlaAspIleGluLeuTyrValLeu	161
Db	128983	ACGCGGTTTTGGAGCGGAATGCGACGCGCAAGCAAAAGGCGGACGTGGGTGTG	129042
Qy	162	GluLeuSerSerPheGlnLeuGluThrCysAspArgLeuAsnAlaGluValAlaThrVal	181
Db	129043	GAGCTTCCAGCTTCCAACTGGAAACACCCGAAAGCCGTGCGTCCGACTCGCGCAGCGTG	129102
Qy	182	LeuAsnValSerGluAspHisMetAspArgTyrAspGlyMetAlaAspTyrHisLeuAla	201
Db	129103	CTGAACATTCCGAAGACCATCTCGACCGCTACGACGACTTGCTCGACTTCGCGATACC	129162
Qy	202	LysHisArgIlePheArgGlyAlaArgGlnValValValAsnArgAlaAspAlaLeuThr	221
Db	129163	AAAGCCAAAGATTTTCCGTGGCGACGGCTGCAGGTTTTGAATCGCGACGATCGCTG---	129219
Qy	222	ArgProLeuIleAlaAspThrValProCysTrpSerPheGlyLeuAsnLysProAspPhe	241
Db	129220	-----TGCCGCGGATGAAGCGTCCGGGCGCGAGGTA	129252
Qy	242	LysAlaPheGlyLeuIleGluGluAspGlyGlnLysTrpLeuAlaPheGlnPhe-----	259
Db	129253	AAATGGTTTCTGTG---GAACACGAAGCTGATTTCTGTTGGAAACGGCAGACAGCGCGC	129309
Qy	260	-----AspLysLeuLeuProValGlyGluLeuLysIleArgGlyAlaHis	274
Db	129310	CTGAACAAGCGCAATGAAGATTGATTGTCAGCAAGACATTCGTTGCAAGPProPheAsp	129369
Qy	275	AsnTyrSerAsnAlaLeuAlaAlaLeuGlyHisAlaValGlyLeuProPheAsp	294
Db	129370	AACGCCCTAACGTCATGCTGCGGTGCTTGTGTGAGGCCATCGGTTGTGCGCGGAA	129429
Qy	295	AlaMetLeuGlyAlaLeuLysAlaPheSerGlyLeuAlaHisArgCysGlnTrpValArg	314
Db	129430	GCATTGCTCGAACACGTCAAACACCTCCAAAGCGCTGCCGCGCAGCGGTGAAATAATCGC	129489
Qy	315	GluArgGlnGlyValSerTyrTyrAspAspSerLysAlaThrAsnValGlyAlaLeu	334
Db	129490	GAGAAAACGGCTGTGTTTATCGACGACACAAAGCAGCATGTGCGCGCGACTGCC	129549
Qy	335	AlaAlaTleGluGlyLeuGlyLysLeuValLeuLeuAlaGlyGly	354
Db	129550	GCCGCGATTGCGGGTTG-----CAAAATCCGCTCTCGTGAATTTGGCGCGC	129597



QY 355 AspGlyLysGlyAlaAspPheHisAspLeuArgGluProValAlaArgPheCysArgAla 374  
 DB 129598 ATGGGTAAAGGAGGAGGACTTCAGCCCTCGCGGATGCAGTGGTAGGCAAGCAAGGC 129657  
 QY 375 ValValLeuLeuGlyArgAspAlaGlyLeuIleAlaGlnAlaLeu---GlyAsnAlaVal 393  
 DB 129658 GTCTTCTGATGTCGATGCGCGCAATCCCGCGGATTTGACGCGCTGCGGCTTG 129717  
 QY 394 ProLeuValArgValAlaThrLeuAspGluAlaValArgGlnAlaAlaGluLeuAlaArg 413  
 DB 129718 AATATGACCGACTGCCCACTTTGGGAGAGCGCTTCAGCGCATATGCCCAAGCCGAA 129777  
 QY 414 GluGlyAspAlaValLeuLeuSerProAlaCysAlaSerLeuAspMetPheTyrAsnPhe 433  
 DB 129778 GCAGCGGATATGTTGTCAGCCCGCTCGCGGAGCTTGTATATGTTCAAGGCTAC 129837  
 QY 434 GluGluArgGlyArgLeuPheAlaLysAlaValGluGluLeu 447  
 DB 129838 GCGCACCGTTCGGAAGTGTATTCGAAGCGTTTAAGGCTTTG 129879

RESULT 13

AAA81490  
 ID AAA81490 standard; DNA; 1437668 BP.  
 XX  
 AC AAA81490;  
 XX  
 DT 04-DEC-2000 (first entry)  
 XX  
 DE N. meningitidis B full length genome DNA sequence SEQ ID NO:1068.  
 XX  
 KW Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;  
 KW antigen; vaccine; diagnosis; infection; antibacterial; identification;  
 KW Meningococcus B; MenB; ds.  
 XX  
 OS Neisseria meningitidis.  
 XX  
 PN WO200022430-A2.  
 XX  
 PD 20-APR-2000.  
 XX  
 PF 08-OCT-1999; 99WO-US23573.  
 XX  
 PR 09-OCT-1998; 98US-0103794.  
 PR 30-APR-1999; 99US-0132068.  
 XX  
 PR (CHIR ) CHIRON CORP.  
 XX  
 PI Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;  
 PI Masignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;  
 PI Rappuoli R, Pizza M;  
 XX  
 DR WPI; 2000-318079/27.

XX Isolated nucleotide sequences of Neisseria meningitidis which can be  
 PT used in the diagnosis and treatment of N. meningitidis infection and  
 PT other Neisserial infections, for example, N.gonorrhoea -  
 XX  
 PS Claim 7; Page 866-1272; 1760pp; English.

XX The present invention describes methods of obtaining immunogenic  
 CC proteins from Neisseria genomic sequences. AAA81453 to AAA82414  
 CC represent specifically claimed Neisseria meningitidis genomic DNA  
 CC sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent  
 CC Neisseria DNA sequences and their corresponding proteins; AAA81254 to  
 CC AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the  
 CC isolation of Neisseria meningitidis DNA sequences; and AAA81322 to  
 CC AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF  
 CC sequences, which are all used in the exemplification of the present  
 CC invention. The nucleic acid sequences, protein sequences, and antibodies  
 CC against them, can be used in the manufacture of a composition. The  
 CC composition can be used as a medicament (or in the manufacture of a  
 CC medicament) for treating, preventing or diagnosing infection due to

CC Neisserial bacteria. For example, some of the identified proteins could  
 CC be components of vaccines against Meningococcus B; against all serotypes;  
 CC and/or against all pathogenic Neisseriae. Identification of sequences  
 CC from the bacterium will also facilitate production of biological probes,  
 CC particularly organism-specific probes. Attempts to make efficacious  
 CC Meningococcus B vaccines have failed mainly due to antigen tolerance.  
 CC Multivalent vaccines have also been tried but none have successfully  
 CC overcome antigenic variability. The provision of further, complete  
 CC sequences may provide an opportunity to identify secreted or surface  
 CC exposed proteins that may be presumed targets for the immune system and  
 CC which are not antigenically variable or at least more conserved than  
 CC other more variable regions.

XX  
 SQ Sequence 1437668 BP; 344338 A; 353206 C; 385074 G; 355045 T; 5 other;

Alignment Scores:  
 Pred. No.: 1-91e-53 Length: 1437668  
 Score: 704.00 Matches: 173  
 Percent Similarity: 53.52% Conservativeness: 70  
 Best Local Similarity: 38.11% Mismatches: 177  
 Query Match: 31.15% Indels: 34  
 DB: 21 Gaps: 8

US-09-701-229-2 (1-448) x AAA81490 (1-1437668)

QY 11 IleValValGlyLeuGlyLysSerGlyMetSerLeuValArgTyrLeuAlaArgGly 30  
 DB 428569 CTGTCGCGGACTCGCGGTATTCATGATGCTACCTACGCAAAACGGC 428628  
 QY 31 LeuProPheAlaValValAspThrArgGluAsnProGluLeuAlaThrLeuArgAla 50  
 DB 428629 GCGAGGTTCTCGTATGATGCGGAGCTGAAGCCGAGCGGTGTCGCAATCGGTAA 428688  
 QY 51 GlnTyrProGlnValGluValArgCysGlyGluLeuAspAlaGluPheLeuCysSerAla 70  
 DB 428689 ATGTTGACGGTGTGTTTACACGGCGCTGGAAGATCGCTGGACACGGTTTC 428748  
 QY 71 ArgGluLeuTyrValSerProGlyLeuSerLeuValArgThrProAlaLeuGlnAlaAla 90  
 DB 428749 GATATTCGTGCTCAGTCCGCGCATCAGCGAGCGGATATGAGCGCTTCAAG 428808  
 QY 91 AlaLysGlyValArgIleSerGlyAspIleAspLeuPheAla 105  
 DB 428809 CAAACGCGGAGCGGCTGTTGGCGGACATCGATGCTGCGGAGACATGTGAACCGCGG 428868  
 QY 106 GluAlaLysAlaProIleValAlaIleThrGlySerAsnAlaLysSerThrValThrThr 125  
 DB 428869 GACGACAAAG-----GTAATTCGATACCGGACGACGCAACCAACCGGTAAACG 428922  
 QY 126 LeuValGlyGluMetAlaValAlaAlaAspLysArgValAlaValGlyGlyAsnLeuGly 145  
 DB 428923 CTGTCGCGCTATCTCTATCAAGTCGCGGTGATACCGTATCGCGGCAATATCGGC 428982  
 QY 146 ThrProAlaLeuAspLeu-----LeuAlaAspAspIleGluLeuTyrValLeu 161  
 DB 428983 ACGCCGCTTTTGGAGCGGAAATGCGAGCGGAGGCAAAAGCGCGGTGTGTTG 429042  
 QY 162 GluLeuSerSerPheGlnLeuGluThrCysAspArgLeuAsnAlaGluValAlaThrVal 181  
 DB 429043 GAGCTTCCAGCTTCCAACTGGAAACACCCGAAAGCTTCGCTCGGCTGCGGCGACGGTG 429102  
 QY 182 LeuAsnValSerGluAspHisMetAspArgTyrAspGlyMetAlaAspTyrHisLeuAla 201  
 DB 429103 CTGAACATTTCCGAGACCATCTCGACCGGTACGACGACATGTCGACTATGCGCATACC 429162  
 QY 202 LysHisArgIlePheArgGlyAlaArgGlnValValValValValValValValValVal 221  
 DB 429163 AAAGCCCAAGATTTTCCGTCGCGGAGCGGTGTCAGGTTTGAATGCGGACGATCGCTTC 429219  
 QY 222 ArgProLeuAlaAlaAspThrValProCysTrpSerPheGlyLeuAsnLysProAspPhe 241  
 DB 429220 -----TGCGCGCGGATGAAGCGTGCAGCGTGCAGCGGCGGAGGTA 429252

QY 242 LysAlaPheGlyLeuLeuGluAspGlyGlnLysTrpLeuAlaPheGlnPhe----- 259  
 DB 429253 AAATGGTTTTCGTTG---GAACACGAAGCTGATTTCTGGTGGCAACGCGAGCGCGC 429309  
 QY 260 -----AspLysLeuLeuProValGlyGluLeuLysLeuLeuGlyAlaHis 274  
 DB 429310 CTGAACACAGGCAATCAAGATTGTGTACCAAGACATTCCTGTTGCAAGCTGTGCAC 429369  
 QY 275 AsnTyrSerAsnAlaLeuAlaLeuAlaLeuGlyHisAlaValGlyLeuLeuProPheAsp 294  
 DB 429370 AAGCGCGCTACGCTAGCTGCGGTGGCTTGTGTAGCCATCGTTGTCGCCGGA 429429  
 QY 295 AlaMeLeuGlyAlaLeuLysAlaPheSerGlyLeuAlaHisArgCysGlnTrpValArg 314  
 DB 429430 GCATTGCTCGAACACCTCAAAACCTTCCAAAGGCTCGCCACCGCTGGGAAAAATCGGC 429489  
 QY 315 GluArgGlnGlyValSerTyrAspAspSerLysAlaThrAsnValGlyAlaAlaLeu 334  
 DB 429490 GAGAAAACGGCTGTGTATTATCGACGACGACGAAAGGCGCAATGTCGGCGGACTGCC 429549  
 QY 335 AlaAlaIleGluGlyLeuGlyAlaAspIleAspGlyLysLeuValLeuAlaGlyGly 354  
 DB 429550 GCGCGATTGCGGTTTG-----CAAAATCCGCTCTTCGTGATTGTCGCGC 429597  
 QY 355 AspGlyLysGlyAlaAspPheHisAspLeuArgGluProValAlaArgPheCysArgAla 374  
 DB 429598 ATGGGTAAAGGCGAGGACTTACGCGCCCTGCGGATGCACTGGTAGGCAAGCAAAAGCG 429657  
 QY 375 ValValLeuLeuGlyArgAspAlaGlyLeuIleAlaGlnAlaLeu---GlyAsnAlaVal 393  
 DB 429658 GTGTTCTGATTGGTTCGATGCGCGCAANTCCCGCGGANTTGGACGGCTGCGGCTTG 429717  
 QY 394 ProLeuValArgValAlaThrLeuAspGluAlaValArgGlnAlaGluLeuAlaArg 413  
 DB 429718 AATATACCGACTGCGCCACTTTGGGAGAAGCGTTGACAGCGCATATGCCCAAGCGAA 429777  
 QY 414 GluGlyAspAlaValLeuSerProAlaCysAlaSerLeuAspMetPheLysAsnPhe 433  
 DB 429778 GCAGGCGATTGTGTCTACGCGCCGCTGCGGAGCTTTGATATGTTCAAAGGCTAC 429837  
 QY 434 GluGluArgGlyArgLeuPheAlaLysAlaValGluGluLeu 447  
 DB 429838 GCGCACCGTTCGGAAGTGTATTATCGAAGCGTTTAAGGCTTTG 429879

## RESULT 14

ABZ41528

ID ABZ41528 standard; DNA; 1335 BP.

XX AC

XX ABZ41528;

XX DT

XX 07-MAR-2003 (first entry)

XX DE N. gonorrhoeae nucleotide sequence SEQ ID 7645.

XX KW Antibacterial; infection; vaccine; gene therapy; gene; ds.

XX XX Neisseria gonorrhoeae.

XX OS

XX PN WO200279243-A2.

XX PD

XX 10-OCT-2002.

XX PF 12-FEB-2002; 2002WO-IB02069.

XX PR 12-FEB-2001; 2001GB-0003424.

XX PA (CHIR-) CHIRON SPA.

XX PI Fontana MR, Piza M, Massignani V, Monaci E;

XX DR WPI; 2003-058415/05.

XX DR P-PSDB; ABP80558.

XX XX

PT New protein from *Neisseria gonorrhoeae*, useful for the manufacture of a  
 medicament for treating or preventing *N. gonorrhoeae* infection -  
 XX  
 PS Disclosure; Page 745-746; 815pp; English.

CC The present invention relates to proteins from *Neisseria gonorrhoeae*.  
 CC Also disclosed are the nucleic acid molecules encoding the proteins and  
 CC antibodies that specifically bind to the proteins. The composition  
 CC comprising the protein, nucleic acid or antibody is useful for the  
 CC manufacture of a medicament for treating or preventing *N. gonorrhoeae*  
 CC infection, this may be in the form of a vaccine or gene therapy.  
 CC Sequences given in records ABZ37706-ABZ42016 represent nucleic acid  
 CC molecules of the invention.

XX Sequence 1335 BP; 303 A; 341 C; 415 G; 276 T; 0 other;

## Alignment Scores:

Pred. No.: 3,78e-57 Length: 1335  
 Score: 701.00 Matches: 173  
 Percent Similarity: 52.88% Conservative: 66  
 Best Local Similarity: 38.27% Mismatches: 183  
 Query Match: 31.02% Indels: 30  
 DB: 25 Gaps: 7

US-09-701-229-2 (1-448) x ABZ41528 (1-1335)

QY 11 IleValValGlyLeuGlyLysSerGlyMetSerLeuValArgTyrLeuAlaArgArgGly 30  
 DB 25 CTCGTGCGCGAGCTTGGCGCAGGTATTCATGATTCATGATTCGCTATCTCGCAAAACGCG 84  
 QY 31 LeuProPheAlaValValAspThrArgGluAsnProProGluLeuAlaThrLeuArgAla 50  
 DB 85 GCGGAGGTTGCGCGCTTATGTCGGAGCTGAAAGCGCGCGTGGCGCAAAATCGGTAAG 144  
 QY 51 GlnTyrProGluValGluValArgCysGlyGluLeuAspAlaGluPheLeuCysSerAla 70  
 DB 145 ATGTGTGACGGCTGTGTGTTTACAGGCTGCTGCTCAAAAGATGCGTGGACACAGGTTTC 204  
 QY 71 ArgGluLeuTyrValSerProGlyLeuSerLeuArgThrProAlaLeuValGlnAlaAla 90  
 DB 205 GATATTCTGCGCTCAGTCCCGCATCAGCGGCGCAGCGCATATCGAGCGGTTCAAG 264  
 QY 91 AlaLysGlyValArgIleSerGlyAspIleAspLeuPheAla-----ArgGluAla 107  
 DB 265 CAAAACGCGCGCGCTGTAGCGGACATCGAATTCGTCGCGGACATTTGTGAACCGCGC 324  
 QY 108 LysAlaProIleValAlaIleThrGlySerAsnAlaLysSerThrValThrThrLeuVal 127  
 DB 325 GCGGACAAGGTGATGCGATTTACCGGACAGCAAGGCAAAACCCAGGATACGAGCTGGTC 384  
 QY 128 GlyGluMetAlaAlaAlaAspLysArgValAlaValGlyLysLeuGlyThrPro 147  
 DB 385 GGCTATCTGTCATCAAGTGGCGGTGGATACCGTACCGTATTCGCGGCAATATCGTACGCGC 444  
 QY 148 AlaLeuAspLeu-----LeuAlaAspAspIleGluLeuTyrValLeuGluLeu 163  
 DB 445 GTTTTGGAGGCAGAAATTGCGAGCGCAAGGCAAAAGCGGACGTGTGGGTGTGGAGCTT 504  
 QY 164 SerSerPheGlnLeuGluThrCysArgLeuAsnAlaGluValAlaThrValLeuAsn 183  
 DB 505 TCCAGCTTCCAACTGGAAAACACCGAAGACCTTGGCGCCGACTGCGCGGAGGTGATGAAC 564  
 QY 184 ValSerGluAspHisMetAspArgTyrAspGlyMetAlaAspTyrHisLeuAlaLysHis 203  
 DB 565 ATTTCCGAAGACCATCTCGACCGCTACGACGATTTGCTCGACTACGCGCATACCAAGGCC 624  
 QY 204 ArgIlePheArgGlyAlaArgGlnValValValAsnArgAlaAspAlaLeuThrArgPro 223  
 DB 625 GAGATTTTCGTCGCGATGGAGTGCAGGTTTTGAAT----- 660  
 QY 224 LeuIleAlaAspThrValProCysTrpSerPheGlyLeuAsnLysProAspPheLysAla 243  
 DB 661 -----GCGGACGACGTGTCTGCGCGCCATGAAGGCGGCGGTGAGTGAACGCG 714



Db 833 ACAGGATCGACGGTAAGACACCAACGACTATGATTGGGAACTTTTGACTGCTGCT 892  
QY 135 AspLysArgValAlaValGlyGlyAsnLeuGlyThrProAlaLeuAspLeu-----Leu 152  
Db 893 GGCAACATGGTCTTTTATCAGGAATATCGGCTATCCAGCTAGTCAAGTGGCCCAACT 952  
QY 153 AlaAspIleGluLeuTyrValLeuGluLeuSerSerPheGlnLeuGluThrCysAsp 172  
Db 953 GCGTCGACAGACACGCTTGTATGGAACCTTCTTCTTCCAACTCATGGGTGTTCAA 1012  
QY 173 ArgLeuAsnAlaGluValAlaThrValLeuAsnValSerGluAspHisMetAspArgTyr 192  
Db 1013 GAATTCATCCAGAGATTCGGGTATTACCAACCTTATGCCAACTCATATCGACTACCAT 1072  
QY 193 AspGlyMetAlaAspTyrHisLeuAlaLysHisArgIle-----PheArgGlyAla 209  
Db 1073 GGGTCATTTTCGAATATAGCAGCAAGTGAATATCCAGAACAAAGATGACAGCAGCT 1132  
QY 210 ArgGlnValValAlaAsnArgAlaAspAlaLeuThrArgProLeuIleAlaAspThrVal 229  
Db 1133 GATTTCTGTCTGTGAACCTTAAATCAAGACTTGACTTCCAGACAGAACCACTGTTGTA 1192  
QY 230 ProCysTrpSerPheGlyLeuAsnLysProAspPheLysAlaPheGlyLeuIleGluGlu 249  
Db 1193 CCATTTTCAACA-----CTGAAAAGTTGAT-----GGACCTTATCTGGAA 1234  
QY 250 AspGlyGlnLysTrpLeuAlaPheGlnPheAspLysLeuLeuProValGlyGluLeuLys 269  
Db 1235 GATGTTCAA-----CTCTACTTCCGTGGTGAAGTAGTCATGCGACGAATGAAATCGGT 1288  
QY 270 IleArgGlyAlaHisAsnTyrSerAsnAlaLeuAlaAlaLeuAlaLeuGlyHisAlaVal 289  
Db 1289 GTTCCAGGTAGCCACAAATGTGAAAATGCCCTGGCAGTATTGCTGTAGCCAAAGCTTCGT 1348  
QY 290 GlyLeuProPheAspAlaMetLeuGlyAlaLeuLysAlaPheSerGlyLeuAlaHisArg 309  
Db 1349 GGTGTGGCAACATCAACCATCAAGGAACCTCTTCAGCCTTCGGTGGTCAACACCGT 1408  
QY 310 CysGlnTrpValArgGluArgGlnGlyValSerTyrTyrAspSerLysAlaThrAsn 329  
Db 1409 CTCAGTTTGTGGATGACATCAAGGGGTAAATCTATACGACAGTAAATCAACTAAT 1468  
QY 330 ValGlyAlaAlaLeuAlaIleGluGlyLeuGlyAlaAspIleAspGlyLysLeuVal 349  
Db 1469 ATCTGGCTACTCAAAAGCCCTGTCAGGATTTGAC-----AACAGCAAGGTCGTC 1519  
QY 350 LeuLeuAlaGlyGlyAspGlyLysGlyAlaAspPheHisAspLeuArgGluProValAla 369  
Db 1520 TTGATTCAGGTGGTTGGACCGTGGCAATGAGTTTGACGAATG---GTGCCAGACATT 1576  
QY 370 ArgPheCysArgAlaValValLeuLeuGlyArgAspAlaGlyLeuIleAlaGlnAlaLeu 389  
Db 1577 ACTGGACTCAGAAGATGTCATCTCGGTCAATCTGCAGAACGTGTCAACGGGCAGCA 1636  
QY 390 GlyAsnAla---ValProLeuValArgValAlaThrLeuAspGluAlaValArgGlnAla 408  
Db 1637 GACAAGGCTGGTGTCTGCTTATGTCGACGACAGATATTGACAGATGCGACCCGCAAGGCC 1696  
QY 409 AlaGluLeuAlaArgGluGlyAspAlaValLeuLeuSerProAlaCysAlaSerLeuAsp 428  
Db 1697 TATGAGCTGCGACTCAGGAGATGTGTTCTTCTTAGTCCCTGCCAATGCCAGCTGGGAT 1756  
QY 429 MetPheLysAsnPheGluGluArgGlyArgLeuPheAlaLysAlaValGluGluLeu 447  
Db 1757 ATGTATGCTAACTTTGAAGTAGCTGGCGACCTCTTTTATCGACACAGTAGCGGAGTTA 1813

Search completed: August 14, 2003, 07:47:54  
Job time : 2055 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n.model

Run on: August 14, 2003, 06:15:45 ; Search time 4239 Seconds  
(without alignments)  
4323.546 Million cell updates/sec

Title: US-09-701-229-2

Perfect score: 2260

Sequence: 1 MSLIASDHFRIVVGLGSGM.....MFKNFEGRGLFAKAVEELA 448

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 4991172

Minimum DB-seq length: 25  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+p2n.model -DRV=xlp  
-O=/cgn2\_1/USPTO\_spool/US09701229/runat\_09082003\_161157\_4380/app\_query.fasta\_1.647  
-DB=GenEmbl -QFMT=fastp -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blos62 -TRANS=human40.cdi -LIST=45  
-DOALIGN=200 -THR\_SCORE=ptc -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=25 -MAXLEN=2000000000  
-USER=US09701229.ecgn\_1\_1.5265,runat\_09082003\_161157\_4380 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6  
-FGAPEXT=7 -FGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl :  
1: gb\_ba :  
2: gb\_hgt :  
3: gb\_in :  
4: gb\_ov :  
5: gb\_pat :  
6: gb\_ph :  
7: gb\_pl :  
8: gb\_pr :  
9: gb\_ro :  
10: gb\_sts :  
11: gb\_sy :  
12: gb\_un :  
13: gb\_vl :  
14: gb\_vl :  
15: em\_ba :  
16: em\_fun :  
17: em\_hum :  
18: em\_in :  
19: em\_mu :  
20: em\_mu :  
21: em\_or :  
22: em\_ov :  
23: em\_pat :  
24: em\_ph :  
25: em\_pl :  
26: em\_ro :  
27: em\_sts :  
28: em\_un :

29: em\_vi :  
30: em\_htg\_hum :  
31: em\_htg\_inv :  
32: em\_htg\_other :  
33: em\_htg\_mus :  
34: em\_htg\_pin :  
35: em\_htg\_rod :  
36: em\_htg\_mam :  
37: em\_htg\_vrt :  
38: em\_sy :  
39: em\_htgo\_hum :  
40: em\_htgo\_mus :  
41: em\_htgo\_other :

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
c 1	2260	100.0	24000	1	AE004856 Pseudomon
2	2249	99.5	5280	1	AY008276 Pseudomon
3	1835	81.2	30195	1	AE016779 Pseudomon
c 4	1829	80.9	311600	1	AE016871 Pseudomon
5	1023	45.3	301727	1	AE016960 Cxiella
6	990	43.8	251050	1	AL627265 Salmonell
7	990	43.8	300169	1	AE016834 Salmonell
8	987	43.7	22348	1	AE008699 Salmonell
9	986	43.6	2608	1	ECMUROY
10	986	43.6	21757	1	AE000118 Escherich
11	986	43.6	28277	6	AX191720 Sequence
12	986	43.6	28277	15	EC2MIN
13	986	43.6	111408	1	EC0110K
14	985.5	43.6	1317	6	AX189058 Sequence
15	984	43.5	12791	1	AE005185 Escherich
16	984	43.5	281530	1	AP002550 Escherich
17	984	43.5	300409	1	AE016755 Escherich
18	978	43.3	11441	1	AE015046 Shigella
19	973	43.1	290029	1	AE016978 Shigella
20	972	43.0	1629	1	ECMURD
c 21	968	42.8	13071	1	AE015855 Shewanell
c 22	965.5	42.7	11726	1	AE013965 Yersinia
c 23	965.5	42.7	201050	1	AJ414143 Yersinia
24	927	41.0	316050	1	BX321859 Nitrosomo
25	908	40.2	21000	1	AB052554 Shewanell
c 26	902	39.9	248308	2	AC073937 Mus muscu
c 27	889.5	39.4	96109	6	AX067460 Sequence
28	879	38.9	11498	1	U32793 Haemophilus
29	879	38.9	11498	6	AX191762 Sequence
30	879	38.9	110000	6	AR274513_11
31	879	38.9	110000	6	AR274513_12
32	875.5	38.7	298900	1	AP005074
c 33	864	38.2	301442	1	AE016798
c 34	859	38.0	188050	1	AL646072 Ralstonia
c 35	844.5	37.4	12232	1	AE006048 Pasteurel
c 36	832	36.8	22201	1	AE004310 Vibrio ch
c 37	757.5	33.5	348600	1	AB063521 Wigglewo
c 38	738	32.7	259498	2	AC020876 Mus muscu
c 39	716	31.7	110000	6	BD061520_2
c 40	716	31.7	347550	1	AP001118 Buchnera
c 41	710	31.4	195767	1	NMA722491
c 42	704	31.2	10629	1	AE002398 Neisseria
43	704	31.2	349980	6	AX044029 Sequence
c 44	694	30.7	310003	1	AE014016 Buchnera
c 45	693.5	30.7	11683	1	AE013119 Thermoana

#### ALIGNMENTS

RESULT 1

AE004856/c	AE004856	24000 bp	DNA	linear.	BCT 19-FEB-2003			
LOCUS	Pseudomonas aeruginosa PA01, section 417 of 529 of the complete genome.							
DEFINITION	Pseudomonas aeruginosa PA01, section 417 of 529 of the complete genome.							
ACCESSION	AE004856	AE004091						
VERSION	AE004856.1	GI:9950633						
KEYWORDS								
SOURCE	Pseudomonas aeruginosa PA01							
ORGANISM	Pseudomonas aeruginosa PA01							
REFERENCE	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas.							
AUTHORS	1 (bases 1 to 24000)							
TITLE	Stover,C.K., Pham,X.-Q.T., Erwin,A.L., Mizoguchi,S.D., Warrenner,P.,							
	Hickey,M.J., Brinkman,F.S.I., Hufnagle,W.O., Kowalik,D.J.,							
JOURNAL	Lagrou,M., Garber,R.L., Goltzy,L., Tolentino,E.,							
	Westbrock-Wadman,S., Yuan,Y., Brody,L.L., Coulter,S.N.,							
MEDLINE	Folger,K.R., Kas,A., Larbig,K., Lim,R.M., Smith,K.A., Spencer,D.H.,							
	Wong,G.K.-S., Wu,Z., and Paulsen,I.T.							
PUBMED	Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen							
REFERENCE	Nature 406 (6799), 959-964 (2000)							
AUTHORS	20437337							
TITLE	10984043							
	2 (bases 1 to 24000)							
JOURNAL	Stover,C.K., Pham,X.-Q.T., Erwin,A.L., Mizoguchi,S.D., Warrenner,P.,							
	Hickey,M.J., Brinkman,F.S.I., Hufnagle,W.O., Kowalik,D.J.,							
MEDLINE	Lagrou,M., Garber,R.L., Goltzy,L., Tolentino,E.,							
	Westbrock-Wadman,S., Yuan,Y., Brody,L.L., Coulter,S.N.,							
PUBMED	Folger,K.R., Kas,A., Larbig,K., Lim,R.M., Smith,K.A., Spencer,D.H.,							
REFERENCE	Wong,G.K.-S., Wu,Z., Paulsen,I.T., Reizer,J., Saier,M.H.,							
AUTHORS	Hancock,R.E.W., Lory,S. and Olson,M.V.							
CONSTRM	Direct Submission							
JOURNAL	Submitted (16-MAY-2000) Department of Medicine and Genetics,							
TITLE	University of Washington Genome Center, University Of Washington,							
	Box 352145, Seattle, WA 98195, USA							
REFERENCE	3 (bases 1 to 24000)							
AUTHORS	Pseudomonas aeruginosa Community Annotation Project (PseudoCAP)							
CONSTRM	Direct Submission							
JOURNAL	Submitted (04-FEB-2003) Department of Molecular Biology and							
TITLE	Biochemistry, Simon Fraser University, 8888 University Dr.,							
	Burnaby, British Columbia V5A 1S6, Canada							
COMMENT	-----							
FEATURES	This represents the February 3, 2003 version of the continually							
	updated, reviewed, Pseudomonas aeruginosa PA01 genome annotation,							
SOURCE	from PseudoCAP (see <a href="http://www.pseudomonas.com">http://www.pseudomonas.com</a> for latest updates							
	and links to alternate annotations). PseudoCAP is coordinated by							
gene	Fiona S.L. Brinkman (Simon Fraser University, Canada) and Robert							
	E.W. Hancock (University of British Columbia, Canada). We welcome							
CDS	submission through <a href="http://www.pseudomonas.com">www.pseudomonas.com</a> of any proposed changes.							
FEATURES	-----							
	'protein name confidence' is used to rate our confidence of the							
Source	accuracy of the protein name.							
gene	Class 1: Function experimentally demonstrated in P. aeruginosa.							
	Class 2: Function of highly similar gene experimentally							
CDS	demonstrated in another organism (and gene context consistent							
FEATURES	in terms of pathways its involved in, if known).							
	Class 3: Function proposed based on presence of conserved amino							





QY 401 LeuaspGluAlaValargGlnAlaAlaGluLeuAlaArgGluGluAlaValLeuLeu 420  
 Db 15065 CTGGACGAGCAGTCCGCGAGCCGCCGAGCTGGCGGCGGATCGCGTGTGTTG 15006  
 QY 421 SerProAlaCysAlaSerLeuAspMetPheLysAsnPheGluGluArgGlyArgLeuPhe 440  
 Db 15005 TCGCGGCGCTGGCGAGCTGGACATGTTCAAGAACTTCGAGAAACGCGGAGCGCTGTC 14946  
 QY 441 AlalysAlaValGluGluLeuAla 448  
 Db 14945 GCCAAGCGGTAGAGGAGCTAGCG 14922

## RESULT 2

## AY008276

## LOCUS

## DEFINITION

## Pseudomonas aeruginosa MurE (murE), MurF (murF), MurG (murG), and MurD (murD) genes, complete cds.

## ACCESSION

## AY008276

## VERSION

## AY008276.2

## KEYWORDS

## SOURCE

## ORGANISM

## Pseudomonas aeruginosa

## Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas.

## REFERENCE

## Azzolina, B.A., Yuan, X., Anderson, M.S. and El-Sherbeini, M.

The cell wall and cell division gene cluster in the *Mra* operon of *Pseudomonas aeruginosa*: cloning, production, and purification of active enzymes

## Protein Expr. Purif. 21 (3), 393-400 (2001)

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## Azzolina, B.A. and Azzolina, B.

Pseudomonas aeruginosa *murA*, *murD*, *murE*, and *murF* genes

## Unpublished

## REFERENCE

## El-Sherbeini, M. and Azzolina, B.

## Direct Submission

## JOURNAL

## REFERENCE

## El-Sherbeini, M. and Azzolina, B.

## Submitted (02-JAN-2001) Biochemistry, Merck Research Laboratories, 126 E. Lincoln Avenue, Rahway, NJ 07065, USA

## JOURNAL

## REMARK

## COMMENT

## FEATURES

## source

## 1..5280

## /organism="Pseudomonas aeruginosa"

## /mol\_type="genomic DNA"

## /db\_xref="taxon:287"

## 1..1464

## /gene="murE"

## /gene="murE"

## /note="UDP-N-acetylmuramoyl-L-alanyl-D-glutamate:meso-2,6-diaminopimelate ligase"

## /codon\_start=1

## /transl\_table=1

## /product="MurE"

## /protein\_id="AAG45235.1"

## /db\_xref="GI:12007716"

## /translation="MPMSLSQLFPQAEEDLLIRELDSHGVRPVDLFLVPGGHQDG RAHIALDTKGTAVAYAEAGELPPSDAPLIAVKGLAQLSAGVAGREFYGEPSGLD LIGVTGNTGTSVQLVAQALDLGERCGIVGLTGTGFGALSCRHWTDPDLAQAT LATIKQAGARAVANEVSHGLDQGRVAALGFDAVFNLSRDHLYHGSWEAYAAAKA KLFAPDRCRLRNLDLDFGRRLAGEEODSELIITYSLTSSAFILYCREASPTASPTAR RSSIAARGLRVLLDFGRFNLNSLAAVALLGLGYPGLDILRTLPQLQGPVGRMRL GGGKGLVVDYATHPDALEKVALELRPHAAALRLCLFGCGGDRDAGKRLPMAIAER LADSVLVTDDNPRTEASAAITADIRKGFAAADKVTFLPFRGEATLAHLIASAAVDDVVL LAGKHEDYQIEDGVRRPFSFDIEQAERLAAWEVPHA"

gene  
 CDS

1457..2833  
 /gene="murF"  
 1457..2833  
 /gene="murF"

/note="recombinant enzyme catalyzed the ATP-dependent addition of D-alanine-D-alanine to the UDP-N-acetylmuramoyl-L-alanine-D-glutamate-m-Dap precursor"

/codon\_start=1

/transl\_table=11

/product="MurF"

/protein\_id="AAG45236.1"

/db\_xref="GI:12007717"

/translation="MLEPLRLSQLTVALDARLIGEDAVFSVSTDSRAIGPQGLFIAL SGPRFDGHDYLAEEAAKGAVALVEREVADAPLPQLLRDTRALRGALNRRKFTG PLAAATSSGKTAVKEMLASILRTQAGDAESLATRGNLNDLGVPLTLQLAOPQRS AVIELGASRIGELIATVELTTPHVAIITNAGTAHVGEFGPEKIVEAKEGILEGLAAD GTAVLNLDKAFDTWKARASGRPLLTFSLDRPDQADFRADLQORDARGCMGRLOGVAG EAQVQLNLLGRHNANALAAAAHAHALGVPLDGVAGLQALQAPYKGRAVAQLTASGLR VIDSYNANPASMIAAIDILSFGSGRTVLVLMGELSGMAEQAHREVSAAAGKVSALYAVGPLMAHAVAFGATGRHFADQASLIGALATLQPTTTILIKGSRSAAMDVKVVAAL CGSSSESH"

2833..3915

/gene="mray"

2833..3915

/gene="mray"

/note="phospho-N-acetylmuramoyl-pentapeptide translocase"

/codon\_start=1

/transl\_table=11

/product="Mray"

/protein\_id="AAG22121.1"

/db\_xref="GI:10719754"

/translation="MLLLAELYLAQFYKFGVQYLTLRGILSVLTALSLSLWLGPMW IRTLQIQIGAVRNDPQSHLSKKGTPMTMGALITLTAISTLWADLSNRYVWVVL VTTLFGAIGWVDYRKVIEKNSGLPSRWKYFWQSVFGIGAAYLYMTATPTPTTL IVPMLKSEIOLGIFFFVVLTVFVLVSSNANLDELGLAIMPVMVAGALGFCYL SGNKAREYLLIPNVPAGELIVFCALVAGLGLFELWENTYPQAVFMGDVGLALGAA LGTIATVIRQEVILFIMGVVFVMTLSVNIQVASFKLTRGRVFRMAPIHDFELKGPW DPRVIRFRIITVILVLIGLATLKL"

3929..5275

/gene="murD"

3929..5275

/gene="murD"

/note="UDP-N-acetylmuramyl-L-alanine:D-glutamate ligase; recombinant enzyme catalyzed the ATP-dependent addition of D-glutamate to the precursor sugar peptide"

/codon\_start=1

/transl\_table=11

/product="MurD"

/protein\_id="AAG45237.1"

/db\_xref="GI:12007718"

/translation="MSLIASDHFRIVVGLKSGMSLVRYLARRGLPFAVVDTRNPPE LATRAQYQVEARCGELDAEFLCARELYVSPGLSLRTPALVQAAGKVRISGIDIL FAREAKAPIVAITGTSNAKSTVTLVGMVAADKRVAVGGNLGTPALDLADLLEYV LELSSFOLETCDRLNAEVATVNLVSEODHMDYDGMADYHLAKHFRFCAROVVNRAD ALTRPLADIVTVCWSPGLNKDPFAKGLIEDGCKWLAFQFKLLIPVSGELIRGAHNY SNALALGHALHVPGLFDMGLKALFAGLHSGALFQVREKQVSTYDDSKATVNGAAL AAIEGLGADIDGLKVLVLAGGDKGADFDLREPVARFCRAVVLGRDAGLTAQALGNA VPLVRVATLDEAVRQAALAREGDVALLSPACSLDMFKNEERGLFAKAVEELA"

BASE COUNT 817 a 1781 c 1759 g 923 t

ORIGIN

Alignment Scores:

Pred. No.: 6,67e-138 Length: 5280

Score: 2249.00 Matches: 446

Percent Similarity: 99.55% Conservative: 0

Best Local Similarity: 99.55% Mismatches: 2

Query Match: 99.51% Indels: 0

DB: 1 Gaps: 0

US-09-701-229-2 (1-448) x AY008276 (1-5280)

QY 1 MetSerLeuIleAlaSerAspHisPheArgTleValValGlyLeuGlyLysSerGlyMet 20

|||||



```
Db 3929 ATGAGCGTGATCGCCCTCCGACCACTTCCCGCATCGTTGTGCGCCTCGCAAGCGCGCATG 3988
Qy 21 SerLeuValArgTyrLeuAlaArgGlyLeuProPheAlaValValAspThrArgGlu 40
Db 3989 TCCTGGTGGCTACCTGGCGCGCGCGGCTTCCCTTTCGCCGTGGTATACCGAGAG 4048
Qy 41 AspProGluLeuAlaThrLeuArgAlaGlnTyrProGlnValGluValArgCysGly 60
Db 4049 AACCCGCGGAGCTGGCTACCTTCGCTGCCAGTATCCGAGGCGTGAAGCGCTTGGCGC 4108
Qy 61 GluLeuAspAlaGluPheLeuCysSerAlaArgGluLeuTyrValSerProGlyLeuSer 80
Db 4109 GAATCGACCGCGAGTTCCTCTGCTCCGCCCGGAATCTATGTACGCCCGGCTTGTGCG 4168
Qy 81 LeuArgThrProAlaLeuValGlnAlaAlaLysGlyValArgIleSerGlyAspIle 100
Db 4169 CTGGCACCCCTCGCTGTGTACAGCGCGCGCGCAAGGCGTGCATCTCCGCTGACATC 4228
Qy 101 AspLeuPheAlaArgGluAlaLysAlaProIleValAlaIleThrGlySerAsnAlaLys 120
Db 4229 GATCTCTTCGCCGCGAGCGCAAGCCCGCATGCTGCCATCACCGGTTCCAAACGCGAAG 4288
Qy 121 SerThrValThrLeuValGlyGluMetAlaValAlaAlaAspLysArgValAlaVal 140
Db 4289 AGCACCTGACACCCCTGGTGGCGAATGCGGTGGCCGCGACAGCGGTGTCGCCGTC 4348
Qy 141 GlyClyAsnLeuGlyThrProAlaLeuAspLeuAlaAspIleGluLeuTyrVal 160
Db 4349 GCGGGCAACCTCGCACCCCGCGCTCGACCTGCTGCGCCGACGACATCGAGCTGAGTGTG 4408
Qy 161 LeuGluLeuSerSerPheGlnLeuGluThrCysAspArgLeuAsnAlaGluValAlaThr 180
Db 4409 TTGGAGCTGTCGAGCTTCCAGCTGGAACCTGCGCATGCGCTCAACGCCGAGTGGCGACC 4468
Qy 181 ValLeuAsnValSerGluAspHisMetAspArgTyrAspGlyMetAlaAspTyrHisLeu 200
Db 4469 GTGCTGACCTGACGAGACCATATGATGCTGACGACGCGATGGCTGACTACCACTG 4528
Qy 201 AlaLysHisArgIlePheArgGlyAlaArgGlnValValValAsnArgAlaAspAlaLeu 220
Db 4529 GCCAAGCAGCGGATCTTCCGCGTTCGCCGCGTGGTGAATCGCGCGCATGCGCTG 4588
Qy 221 ThrArgProLeuIleAlaAspThrValProCysTrpSerPheGlyLeuAsnLysProAsp 240
Db 4589 ACCGACCGCTGATCGCGCATACCGTCCGCTGCTGGTGTGCTGGCCCTGAACAGCGCGAC 4648
Qy 241 PheLysAlaPheGlyLeuIleGluGluAspGlyGlnLysTrpLeuAlaPheGlnPheAsp 260
Db 4649 TTCAAGCTTTCGGCTGTATCGAGGAGACGGCCAGAGTGGCTGGCGTTCAGATTCGCGC 4708
Qy 261 LysLeuLeuProValGlyGluLeuLysIleArgGlyAlaHisAsnTyrSerAsnAlaLeu 280
Db 4709 AAGCTGCTGCGGTTGGCAACTGAAGATCCGTGGCGGCCACCAACTATTTCACAGCGGCTC 4768
Qy 281 AlaAlaLeuAlaLeuGlyHisAlaValGlyLeuProPheAspAlaMetLeuGlyAlaLeu 300
Db 4769 GCGCGCTGCGCTGGGCCCATCGGTGCGCTGCGCTGCGCTGCGCATGCTCGCGCGCTG 4828
Qy 301 LysAlaPheSerGlyLeuAlaHisArgCysGlnTrpValArgGluArgGlnGlyValSer 320
Db 4829 AAGCGCTTTTCGGCGCTTGGCTATCGCTGCGTGGGTGCGGTGCGGAGCGGCGGCGTGGAGC 4888
Qy 321 TyrTyrAspSerLysAlaThrAsnValGlyAlaAlaLeuAlaIleGluGlyLeu 340
Db 4889 TACTACGACCATTCAGAGCCACCAAGCTCGGCGCGCCCTCGCGGCGATCGAGGGGCTG 4948
Qy 341 GlyAlaAspIleAspGlyLysLeuValLeuAlaGlyGlyAspGlyLysGlyAlaAsp 360
Db 4949 GGTGCCGACATCGACGCAAGCTGGTGTCTGCTCGCGCGGAGAGCGCAAGGCGCGCAT 5008
Qy 361 PheHisAspLeuArgGluProValAlaArgPheCysArgAlaValValLeuLeuGlyArg 380
Db 5009 TTCATGACCTGCGGAGCGGCTGCGCGCTTCTGCGGCGGTGGTGGTACTGCTTGGCGCGT 5068
```

---

```
Qy 381 AspAlaGlyLeuIleAlaGlnAlaLeuGlyAsnAlaValProLeuValArgValAlaThr 400
Db 5069 GACGCGGGCTGATTGCCAGGCACCTGGGACACGGGTACCGTGGTGGCGCGTGCACAGC 5128
Qy 401 LeuAspGluAlaValArgGlnAlaAlaGluLeuAlaArgGluGlyAspAlaValLeuLeu 420
Db 5129 CTGGAGGAAGCAGTCCGGCAGCGCGGAGCTGGCGCGGAGGCGATGCGGTGCTGTG 5188
Qy 421 SerProAlaCysAlaSerLeuAspMetPheLysAsnPheGluGluArgGlyArgLeuPhe 440
Db 5189 TCGCCGCGCTGCGGAGCCCTGGACATGTTCAAGAACTTCGAAGAACCGGACCGCTGTC 5248
Qy 441 AlaLysAlaValGluGluLeuAla 448
Db 5249 GCCAAAGCGGTAGAGGAGCTAGCG 5272

RESULT 3
LOCUS AE016779
DEFINITION Pseudomonas putida KT2440 section 6 of 21 of the complete genome.
ACCESSION AE016779
VERSION AE016779.1
KEYWORDS GI:26557023
SOURCE Pseudomonas putida KT2440
ORGANISM Pseudomonas putida KT2440
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE 1 (bases 1 to 301995)
AUTHORS Nelson,K., Paulsen,I., Weinel,C., Dodson,R., Hilbert,H., Fouts,D.,
Gill,S., Pop,M., Martins Dos Santos,V., Holmes,M., Brinkac,L.,
Beanan,M., DeBoy,R., Daugherty,S., Kolonay,J., Madupu,R.,
Nelson,W., White,O., Peterson,J., Khouri,H., Hance,I., Lee,P.,
Holtzapple,E., Scanlan,D., Tran,K., Moazzaz,A., Utterback,T.,
Rizzo,M., Lee,K., Kosack,D., Tran,K., Moazzaz,A., Utterback,T.,
Hohseisel,J., Straetz,M., Heim,S., Kiewitz,C., Eisen,J., Timmis,K.,
Duesterhoft,A., Tummeler,B. and Fraser,C.
Complete genome sequence and comparative analysis of the
metabolically versatile Pseudomonas putida KT2440
Environ. Microbiol. 4 (12), 799-808 (2002)
REFERENCE 2 (bases 1 to 301995)
AUTHORS Nelson,K., Paulsen,I., Weinel,C., Dodson,R., Hilbert,H., Fouts,D.,
Gill,S., Pop,M., Martins Dos Santos,V., Holmes,M., Brinkac,L.,
Beanan,M., DeBoy,R., Daugherty,S., Kolonay,J., Madupu,R.,
Nelson,W., White,O., Peterson,J., Khouri,H., Hance,I., Lee,P.,
Holtzapple,E., Scanlan,D., Tran,K., Moazzaz,A., Utterback,T.,
Rizzo,M., Lee,K., Kosack,D., Tran,K., Moazzaz,A., Utterback,T.,
Hohseisel,J., Straetz,M., Heim,S., Kiewitz,C., Eisen,J., Timmis,K.,
Duesterhoft,A., Tummeler,B. and Fraser,C.
Direct Submission
Submitted (05-NOV-2002) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
FEATURES
source
1. 301995
/organism="Pseudomonas putida KT2440"
/mol_type="genomic DNA"
/strain="KT2440"
/db_xref="taxon:160488"
89. .712
/gene="sspA"
/locus_tag="Pp1320"
89. .712
/gene="sspA"
/locus_tag="Pp1320"
/notes="similar to GP:9864180; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="stringent starvation protein A"
/protein_id="AAG66944.1"
/db_xref="GI:24982777"
/translation="MGATNRACYSDPADHYSRVRVLVLAEGKSVGLIDVDPALHPR
KLAENVPGSVPTLVDRDLALYESTVVMYLEERYPLPMPYPVARGNSRLMHRI
```

```

QRDMCALADTVLDPRSEARTEARKALRESITGVSLFSEFACFMSEDEQSLVDCCLL
PILWRLPVIGIELPQAKPLLDYMERQFAPEFQASLSVERMKL"
693. .1154
/ gene="aspb"
/ locus_tag="pp1321"
CDS
693. .1154
/ gene="aspb"
/ locus_tag="pp1321"
/ note="similar to GB:S50157, GB:X60221, SP:P24539, and
PID:509291; identified by sequence similarity; putative"
/ codon_start=1
/ transl_table=11
/ product="stringent starvation protein B"
/ protein_id="AAN66945.1"
/ db_xref="GI:24982778"
/ translation="MKCAFSEPLMNSRRPYLVRLYEWIVDNCPTPHMLYNAEPKV
QVPFASDQGVILNFSAVKSLHMDNDAVSFEGRFSVAHSLFVPGAILGIYARE
NQGMVFEELEPSSMDGDEDDVQPPDDGPPDDGGQPPRTGRPSLKVK"
complement(1216..1794)
/ locus_tag="pp1322"
CDS
complement(1216..1794)
/ locus_tag="pp1322"
/ codon_start=1
/ transl_table=11
/ product="lipoprotein, putative"
/ protein_id="AAN66946.1"
/ db_xref="GI:24982779"
/ translation="WTPMRLGLMALTCGLSVTGCSSVLTSTNSPIEDDCTRTTIGSK
IDDSLEIKAAVNIISKANPNLDKSHIVVSSYNGIYLLAGOTPRADLKSLEQTSQV
QRVKVHNEQLQWQSSILARNNDALWLTIKTKTQMLTSDAVPSSRIKIVITENGIVYLL
GLVTQEA NSTAVVQGVSGQIKVLFYID"
complement(1791..2378)
/ gene="gmhA"
CDS
complement(1791..2378)
/ locus_tag="pp1323"
/ gene="gmhA"
/ locus_tag="pp1323"
/ note="identified by match to PFAM protein family HMM
PF01380"
/ codon_start=1
/ transl_table=11
/ product="phosphoheptose isomerase"
/ protein_id="AAN66947.1"
/ db_xref="GI:24982780"
/ translation="MQSRIIRLFESIETQOAMDILAPHIEQASLVNVLNNEGKM
LACNGGAGAQHFSSELLNRFERESLPALATLTDSSLTSTANDYSYNEVFSKO
IRALQPGDVLIAISTGNSANVIOAIOAHDRMIVVALTGRDGGNASLLLPEDVE
complement(2471..2845)
/ locus_tag="pp1324"
CDS
complement(2471..2845)
/ locus_tag="pp1324"
/ note="identified by match to TIGR protein family HMM
TIGR00252"
/ codon_start=1
/ transl_table=11
/ product="conserved hypothetical protein TIGR00252"
/ protein_id="AAN66948.1"
/ db_xref="GI:24982782"
/ translation="MMAASPTRAGQAEATOLEVLOGGLQALLARNWRCKGGELDLV
MLDADTVFVSRYRLHAGFGGALDSIDGRQKRLVLAATLFLQKEAHGWNPCRFDV
VALQSHAGAPLQWLKNAFC"
complement(2842..4659)
/ locus_tag="pp1325"
CDS
complement(2842..4659)
/ locus_tag="pp1325"
/ codon_start=1
/ transl_table=11
/ product="lipoprotein, putative"
/ protein_id="AAN66949.1"
/ db_xref="GI:24982783"
/ translation="WIACLRLLTALCLAAALAACASSPSSSLGELPPTPDASIEQLLE
KAASSKSDAALLRLSNADLAYKQDFPRAARILEQVPLDTLPAQOVFASTLAEL

```

```

AMSRNQPKAALTALAHPSLQRLAEPLADLQARTYSVHAAALEADQALAAQAQRILLA
PLLGQAASNTNDALWALVASLPABQLQOQNTDQTLAGWTSLAFVKSAGTLEQOQAA
IEAVYKHDPHAAQOLPLATLTKELASQPLTKALLPQEGPLAGYARALRDEQMA
AHFQAQSQGPAPAVQVEDSSRIGSLDDFYRQAQAAGVOLVIGPLEKFLYQLAKPQ
LPITTLALNYADAGQKAPQOLFQGLAAEDAREVARARRADMGWRAVALVPSGEMGD
RVLAARODWEGNGTLLAERIAQOPVALAQIADFLQRQSEGRKSLQSTVGSGSIA
AQPSSRRQIDIFTLASTPQQAQIKPTLNFQVAGDVPVYATSNLYSASGDILOYNDMN
GIRFCETPMLDITNSLRQOVVOQWPAAGSISRLYANGVDAYS LAPRLGQALKALPDN
RVILGSGLSINANQVRERQLPWAEEFSGQVKRLPDTVR"
4837. .5721
/ locus_tag="pp1326"
CDS
4837. .5721
/ locus_tag="pp1326"
/ note="identified by match to PFAM protein family HMM
PF00590"
/ codon_start=1
/ transl_table=11
/ product="tetracycline methylase family protein"
/ protein_id="AAN66950.1"
/ db_xref="GI:24982784"
/ translation="MCAVTDVAGASKSTVGTLYVVTATPISGLNDMSARALKVLADVAL
TAAEDTRHSIRLLQHFQIDTPLAACHEHNERDEGRFLSKLLAGENVALVSDAGTPLI
SDPGTHLVQRQAAGVSVVPFGACALIAAAGLPSDRIFEGFLPAKQARRARL
EQVREPTLIFYEAPHRILECLEDMEAVFGDRLALLARELTFTFELGLPLAELR
AFVAGDSNQQRGECVVLVGVWSAPEGEQAISAEARVLDLLLAELPLKRAALAAEIT
GVKNLLYQLALEKQKAB"
complement(5744..6164)
/ locus_tag="pp1327"
/ note="This region contains a gene with one or more
premature stops or frameshifts, and is not the result of a
sequencing artifact; conserved hypothetical protein,
degenerate; similar to GB:X16944, and SP:P15488;
identified by sequence similarity; putative"
6455. 6928
/ locus_tag="pp1328"
CDS
6455. .6928
/ locus_tag="pp1328"
/ note="identified by match to PFAM protein family HMM
PF02381"
/ codon_start=1
/ transl_table=11
/ product="conserved hypothetical protein TIGR00242"
/ protein_id="AAN66951.1"
/ db_xref="GI:24982786"
/ translation="MWGITVFRGANAVSLDAKGLAMPSPRYRDELDSRCNGQLIVTI
DAVPCLCVYPLDWEQIEAKRLALPSRENNRRLQRLIGNVLELDGSGRFLVPP
RLREYAKLDKAMLVQLNKFQLEDWEDANNAVSAADLAAIQPGAMPDELRLIL"
6925. .7872
/ locus_tag="pp1329"
CDS
6925. .7872
/ locus_tag="pp1329"
/ note="similar to GP:9950652; identified by sequence
similarity; putative"
/ codon_start=1
/ transl_table=11
/ product="conserved hypothetical protein TIGR00006"
/ protein_id="AAN66952.1"
/ db_xref="GI:24982787"
/ translation="MTIDSGFNHTVLLDEAVEALALRADGCYLDGTGFRGHSRLIL
SKLPQRLGLGFDKDPQAIATGQALAAEDGRFVIVQSFAGELAEVAARGLHKVSGV
LLDGVSSPQLDDPERGFSFLNDGPLDMRNDPGVSAAEIATAPVEETIARVKEYGV
EERFAGMARAVVERREKQPTFTADLAELVLKVANPAWEKGNPATFADQLRLTHVN
ELGDLEAGLAALDALEVGGRLAYISFHSLEDRIVLKLFRMRKLVKRGADNLPRLNPVQH
KVFEFKIKLIGKAQFASEALKANPRSRSAVMVAEKL"
7869. 8162
/ gene="ftsL"
CDS
7869. .8162
/ locus_tag="pp1330"
/ gene="ftsL"
/ locus_tag="pp1330"
/ note="similar to GB:M11186, GB:M62611, SP:P01178,
PID:L89411, PID:386991, and PID:825699; identified by
sequence similarity; putative"

```

.....

AE016871	311600 bp	DNA	linear	BCT 05-MAR-2003
LOCUS				
DEFINITION	Pseudomonas syringae pv. tomato str. DC3000 section 16 of 21 of the complete genome.			
ACCESSION	AE016871	AE016853		
VERSION	AE016871.1	GI:28854552		
KEYWORDS				
SOURCE	Pseudomonas syringae pv. tomato str. DC3000			
ORGANISM	Pseudomonas syringae pv. tomato str. DC3000			
REFERENCE	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas.			
AUTHORS	1 (bases 1 to 311600)			
	Buell, R., Joardar, V., Khouri, H., Fedorova, N., Tran, B., Russell, D., Berry, K., Utterback, T., Van Aken, S., Feldblyum, T., Gwinn, M., Dodson, R., DeBoy, R., Durkin, A., Kolonay, J., Madupu, R., Daugherty, S., Brinkac, L., Beanan, M., Haft, D., Selengut, J., Nelson, W., Davidesen, T., White, O., Fraser, C. and Collmer, A.			
TITLE	Complete Sequence of Pseudomonas syringae			
JOURNAL	Unpublished			
REFERENCE	2 (bases 1 to 311600)			
AUTHORS	Buell, R., Joardar, V., Khouri, H., Fedorova, N., Tran, B., Russell, D., Berry, K., Utterback, T., Van Aken, S., Feldblyum, T., Gwinn, M., Dodson, R., DeBoy, R., Durkin, A., Kolonay, J., Madupu, R., Daugherty, S., Brinkac, L., Beanan, M., Haft, D., Selengut, J., Nelson, W., Davidesen, T., White, O., Fraser, C. and Collmer, A.			
TITLE	Direct Submission			
JOURNAL	Submitted (03-MAR-2003) The Institute for Genomic Research, 9712			
FEATURES	Medical Center Dr, Rockville, MD 20850, USA			
source	Location/Qualifiers			
	1. 311600			
	/organism="Pseudomonas syringae pv. tomato str. DC3000"			
	/mol_type="genomic DNA"			
	/strain="DC3000"			
	/db_xref="taxon:223283"			
	/note="pathovar: tomato"			
gene	131..601			
	/gene="bfr"			
CDS	locus_tag="PSPT04160"			
	131..601			
	/gene="bfr"			
	locus_tag="PSPT04160"			
	/note="similar to SP:P22759, and SP:P22759; identified by			

gene	CDS	sequence similarity; putative"	sequence similarity; putative"
		/codon_start=1	/codon_start=1
gene	CDS	/transl_table=1	/transl_table=1
		/product="bacterioferritin"	/product="bacterioferritin"
gene	CDS	/protein_id="AAO57616.1"	/protein_id="AAO57616.1"
		/db_xref="GI:28854553"	/db_xref="GI:28854553"
gene	CDS	/translation="MKGDITVTOHNLKIGLNELVAINOYFLHARMYEDWGLKKLGHE YKESHAMKYADKLKRTFLLEGPNLQDLKILIGHTKEMLECDLKIEQKGLADLK AAIAHFETVDFGSRLELLEPKEEQIDWLETQLGLTKIVGNIENLQSQMGD"	/translation="MKGDITVTOHNLKIGLNELVAINOYFLHARMYEDWGLKKLGHE YKESHAMKYADKLKRTFLLEGPNLQDLKILIGHTKEMLECDLKIEQKGLADLK AAIAHFETVDFGSRLELLEPKEEQIDWLETQLGLTKIVGNIENLQSQMGD"
		complement(689..1015)	complement(689..1015)
gene	CDS	/locus_tag="PSPT04161"	/locus_tag="PSPT04161"
		complement(689..1015)	complement(689..1015)
gene	CDS	/locus_tag="PSPT04161"	/locus_tag="PSPT04161"
		/note="identified by match to TIGR protein family HMM TIGR00365"	/note="identified by match to TIGR protein family HMM TIGR00365"
gene	CDS	/codon_start=1	/codon_start=1
		/transl_table=1	/transl_table=1
gene	CDS	/product="glutaredoxin-related protein"	/product="glutaredoxin-related protein"
		/protein_id="AAO57617.1"	/protein_id="AAO57617.1"
gene	CDS	/db_xref="GI:28854554"	/db_xref="GI:28854554"
		/translation="MDIETIKEQIASNTILYMKGAPNAPQCFSKAKASQALMACGE KFAYVDILQNPRIANLPKRYANWTFPQLWVAGELVGGSDIITEMNADGSLQTLVKEA SAAKAE"	/translation="MDIETIKEQIASNTILYMKGAPNAPQCFSKAKASQALMACGE KFAYVDILQNPRIANLPKRYANWTFPQLWVAGELVGGSDIITEMNADGSLQTLVKEA SAAKAE"
gene	CDS	complement(1104..3302)	complement(1104..3302)
		/locus_tag="PSPT04162"	/locus_tag="PSPT04162"
gene	CDS	complement(1104..3302)	complement(1104..3302)
		/locus_tag="PSPT04162"	/locus_tag="PSPT04162"
gene	CDS	/note="similar to GP:5441785; identified by sequence similarity; putative"	/note="similar to GP:5441785; identified by sequence similarity; putative"
		/codon_start=1	/codon_start=1
gene	CDS	/transl_table=1	/transl_table=1
		/product="oxidoreductase, molybdopterin-binding"	/product="oxidoreductase, molybdopterin-binding"
gene	CDS	/protein_id="AAO57618.1"	/protein_id="AAO57618.1"
		/db_xref="GI:28854555"	/db_xref="GI:28854555"
gene	CDS	/translation="MARVIRHSHCSAHRITAMTRPEAAMTRTOHLRACHLCEAICG LAIEWTEPGAAPKITSIKGDPDTSFRGHICPKAVAQIDQNDPDRQPMLEPTGDO WPIEQAFDLVAFELVIAIQOQHONAVAYQGNPSVHNYGLMTHSNYFLGLLKTNRN RSATSVQQLPHLISFLMYGMLLPIDIDHTDFMLILGNPLASNGSIMVDPVE KRKAIAQGGKLVDPDRRSETAIAIOQHVFPVPGDAALLFLGLTLFEEGLPRES HIPVGLDHRHSIAGFSAEAMSPRCIAAQAIRQARDFAGADTALCYRGVSTQA FTCLCHWLQNLINVTGNLDVGGTLCTEPAVDLVSTSGGHFNWQSRVSGLPYGG ELPSALAEMLVEGQGVRAVTVAGNPVLTGPNRGDQDALEGLFMLSIDLVLAA TTRHADLILPSTALENDHYDTENTLAVRNVTRFNAIFDKPEGALHDMWFVGLAA SFAAKAQLKPTVPQAQIMDLVLRAGRYGAASPHNLSLETLASPHMGDGLALAPNL TRLTANGRIQAPEVIMADLVRFADAGPRFGKAGOLLIGRRHVRSSNMWHNY HLRVKGPRLQMLHPDLSLHRLGSDGQGVRSRVIEVEVLGSLDMQPGVWSLPH GWGHSRGYKMEIARNPQVGSANDLTERQDLSGNAALNVPVQVASC"	/translation="MARVIRHSHCSAHRITAMTRPEAAMTRTOHLRACHLCEAICG LAIEWTEPGAAPKITSIKGDPDTSFRGHICPKAVAQIDQNDPDRQPMLEPTGDO WPIEQAFDLVAFELVIAIQOQHONAVAYQGNPSVHNYGLMTHSNYFLGLLKTNRN RSATSVQQLPHLISFLMYGMLLPIDIDHTDFMLILGNPLASNGSIMVDPVE KRKAIAQGGKLVDPDRRSETAIAIOQHVFPVPGDAALLFLGLTLFEEGLPRES HIPVGLDHRHSIAGFSAEAMSPRCIAAQAIRQARDFAGADTALCYRGVSTQA FTCLCHWLQNLINVTGNLDVGGTLCTEPAVDLVSTSGGHFNWQSRVSGLPYGG ELPSALAEMLVEGQGVRAVTVAGNPVLTGPNRGDQDALEGLFMLSIDLVLAA TTRHADLILPSTALENDHYDTENTLAVRNVTRFNAIFDKPEGALHDMWFVGLAA SFAAKAQLKPTVPQAQIMDLVLRAGRYGAASPHNLSLETLASPHMGDGLALAPNL TRLTANGRIQAPEVIMADLVRFADAGPRFGKAGOLLIGRRHVRSSNMWHNY HLRVKGPRLQMLHPDLSLHRLGSDGQGVRSRVIEVEVLGSLDMQPGVWSLPH GWGHSRGYKMEIARNPQVGSANDLTERQDLSGNAALNVPVQVASC"
		complement(3357..3650)	complement(3357..3650)
gene	CDS	/locus_tag="PSPT04163"	/locus_tag="PSPT04163"
		complement(3357..3650)	complement(3357..3650)
gene	CDS	/locus_tag="PSPT04163"	/locus_tag="PSPT04163"
		/note="identified by Glimmer2; putative"	/note="identified by Glimmer2; putative"
gene	CDS	/codon_start=1	/codon_start=1
		/transl_table=1	/transl_table=1
gene	CDS	/product="conserved hypothetical protein"	/product="conserved hypothetical protein"
		/protein_id="AAO57619.1"	/protein_id="AAO57619.1"
gene	CDS	/db_xref="GI:28854556"	/db_xref="GI:28854556"
		/translation="MDSKEHQTELLGLTARTLTHLTASMTSMSEFLLRSEDEVTRNA GRMDTRATSSGLDEHWRLLGLDTGIHMAQEQTETVTEIQLQRKIPTIGG"	/translation="MDSKEHQTELLGLTARTLTHLTASMTSMSEFLLRSEDEVTRNA GRMDTRATSSGLDEHWRLLGLDTGIHMAQEQTETVTEIQLQRKIPTIGG"
gene	CDS	4100..5020	4100..5020
		/gene="argf"	/gene="argf"
gene	CDS	/locus_tag="PSPT04164"	/locus_tag="PSPT04164"
		4100..5020	4100..5020
gene	CDS	/gene="argf"	/gene="argf"
		/locus_tag="PSPT04164"	/locus_tag="PSPT04164"
gene	CDS	/note="identified by match to PFAM protein family HMM PF02729"	/note="identified by match to PFAM protein family HMM PF02729"
		/codon_start=1	/codon_start=1
gene	CDS	/transl_table=1	/transl_table=1
		/product="ornithine carbamoyltransferase"	/product="ornithine carbamoyltransferase"
gene	CDS	/protein_id="AAO57620.1"	/protein_id="AAO57620.1"
		/db_xref="GI:28854557"	/db_xref="GI:28854557"
gene	CDS	/translation="NNARHFLSMNDYTPDELLGLIRRCVELKDLNRGVLFEPLKNRV	/translation="NNARHFLSMNDYTPDELLGLIRRCVELKDLNRGVLFEPLKNRV

LGMIFEKSTRRLSFEAGMIQOLGQAIIFLSHRDTQLGRGEPIADSAKVISRMLDAVM  
 IRTYAHNLTFAANSRVPVINGLSDDLHPCLADLMQTFLEHRSIGKTKTVAMICDG  
 NMNCSYIEAAIOFQOLRVACPAGYEPNPELALAGERTVVRQPKAAVAGAHVSTG  
 DWTSWGOEETAREKALFAPQVTRALLDLADKDVLFMHCLPAHGRBEISVDLLDDA  
 RSVADQARENRLHAQKALLEFLVAFQPA"  
 5026..6087  
 /locus\_tag="PSPT04165"  
 5026..6087  
 /locus\_tag="PSPT04165"  
 /note="similar to GB:248042, SP:Q14444, and PID:662594;  
 identified by sequence similarity; putative"  
 /codon\_start=1  
 /transl\_table=1  
 /product="ABC transporter, ATP-binding protein"  
 /protein\_id="AAO57621.1"  
 /db\_xref="GI:28854558"  
 /translation="MLLNRLACGYQNRVVQDLNLHNLNAGDICGLAGSSCGKTKTT  
 LRAIAGFEIHAGEITLAGDVISRPGWTALPEKRRIGWQFYALFPHLSVADNIAFG  
 IRQPRLDQVIGELLEVLNGLKRGYRPHLESGGQQQRVALARALAPLPQLLLDEPF  
 SNLDELRRRISHEVRDILKARGTSAILVTHDDEAFVSDHVGVEKGRLEQMDTPY  
 NLYHEPQDPFVASFYIGQYFIRGQMIPEVSHTELGVLRGNRAYPCVRGSADVLLRP  
 DDIVVAPSDILKARVIGRTFKGASTLYRLQLPTGSQLEAIPPSHADHRQGEVGRVA  
 ADHLVFAAGSVAAVAAAG"  
 complement(6096..6566)  
 /locus\_tag="PSPT04166"  
 complement(6096..6566)  
 /locus\_tag="PSPT04166"  
 /note="similar to SP:P21484; identified by sequence  
 similarity; putative"  
 /codon\_start=1  
 /transl\_table=1  
 /product="ybak/ebsc protein"  
 /protein\_id="AAO57622.1"  
 /db\_xref="GI:28854559"  
 /translation="MTPALDILLKNAEHRTHSYHDPKAAASYGLEAAEKLGLDAARY  
 EKTLLASKEGELLVAVVVGTVTDLKALANNAKKTETADPAAPASTQYLGGIS  
 PLGKKRRTIDETAQLFESIYSAGRRGLEVELTAAVLAETQARPAFIGRE"  
 6918..7775  
 /gene="glpF"  
 /locus\_tag="PSPT04167"  
 6918..7775  
 /gene="glpF"  
 /locus\_tag="PSPT04167"  
 /note="similar to SP:pl1244; identified by sequence  
 similarity; putative"  
 /codon\_start=1  
 /transl\_table=1  
 /product="glycerol uptake facilitator protein"  
 /protein\_id="AAO57623.1"  
 /db\_xref="GI:28854560"  
 /translation="MTIALKQPTLTGQCVAFELGTALMIFFGTCVAAKLVAGATFGL  
 WEICILMGVMSGIYLSAGHNLNPAVSLASFAGFEKRLKLPFYISAIAGAFG  
 GAGVILLYLSLFFDFEHAHHIIRGSEQLASVSTYPNPASVGOAFLVEVITT  
 ILMGVIMAGDDNGLPRGALPILGLLVAVIGSAMGLTGFAWNPARDGPKLMTF  
 FAGGEIATGGRDIPYLPVPIFAPILGACIAGACYRALIARHLHPSTVTVENEKAPV  
 VRGRVQASS"  
 7805..9310  
 /gene="glpK"  
 /locus\_tag="PSPT04168"  
 7805..9310  
 /gene="glpK"  
 /locus\_tag="PSPT04168"  
 /note="identified by match to PFAM protein family HMM  
 PF02782"  
 /codon\_start=1  
 /transl\_table=1  
 /product="glycerol kinase"  
 /protein\_id="AAO57624.1"  
 /db\_xref="GI:28854561"  
 /translation="MTDTKNKNIYIALDQGTSSRAIIFDRDANVNVSTAQSEFVQHYP  
 QAGVWEHDPMEIFATQACMTKALQAQADLHDDQIAAIGITNORETTVWERDTRPIY  
 NAIWQCRSRSTEICQOLKRCGLGEYIKDTTGLVDLPDFSGSKVKWILDHVSGSRAR  
 KGLMEFTIDWLKFTGGKGVHVTDTYNASRTMLFNIHTLENDQRLDLDVJDFREIL

```
PEVKSSEYVGHKSGIPITAGIDQQAALFGQMVEPQAKNTYGTGCFLLMNTGKK
AVKSAHGLYTTTCGPRGEVVALEGAIVFGSTQVWLDELKLIINDALDTYFAGKV
KDSNGVLYPAFTGLGAPYDVAARGALFGLRGVVDHIIRAAELSYAQTROVLDA
MOQDSEERKLSLRVDGGA VANNFLMQFQADILGTHVERPOMRETTALGRAFLAIG
FWSLDELKNAKAVIERVFPFACEEAHRKLYAGWKQKAVARTDNEPHENE"
9421..10176
/genes="g1pR"
/locus_tag="PSP04169"
9421..10176
/genes="g1pR"
/locus_tag="PSP04169"
/notes="similar to SP:Q51391; identified by sequence
similarity; putative"

Alignment Scores:
Pred. No.: 2 86e-108 Length: 311600
Score: 1829.00 Matches: 353
Percent Similarity: 89.29% Conservative: 47
Best Local Similarity: 78.79% Mismatches: 48
Query Match: 80.93% Indels: 0
DB: 0 Gaps: 0

US-09-701-229-2 (1-448) x AE016871 (1-311600)
QY 1 MetSerLeuLeuAlaSerAspHisPheArgIleValValcIlyLeuGlyLysSerGlyMet 20
Db 291852 GTGTCCTGATGCTTTCCGACCGCTTCCGATTTGTTGGGGCTTGGCAAGAGTGGCATG 291793
QY 21 SerLeuValArgTyrLeuAlaArgArgGlyLeuProPheAlaValValAspThrArgGlu 40
Db 291792 TCCTCGTGTGCTTCTGCGCAACCGGGGCGTCTCTTTGCTGTGCGCCACACGCGGGAA 291733
QY 41 AspProGluLeuAlaThrLeuArgAlaGlnTyrProGlnValGluValArgCysGly 60
Db 291732 AATCTCTCTGAGCTGGCGATTGCGTGTGACTATCCTCAGGTGGAAGTGGGTGGC 291673
QY 61 GluLeuAspAlaGluPheLeuCysSerAlaArgGluLeuTyrValSerProGlyLeuSer 80
Db 291672 GAATCGACGTGGACTCTCTGCGCGGACGAGCTTATGTCAGCCCTGCTGGCC 291613
QY 81 LeuArgThrProAlaLeuValGlnAlaAlaLysGlyValArgIleSerGlyAspIle 100
Db 291612 CTGGCCACACCGGCTTGCAGCAGCGCATGTCAGCGGCGCCAGCTGTCGGCGACATC 291553
QY 101 AspLeuPheAlaArgGluAlaLysAlaProIleValAlaIleThrGlySerAsnAlaLys 120
Db 291552 GAATGTTGCGCGTTACGCCAAGCAGCGGTGATTCGATTCACCGTTCCACAGCGAAA 291493
QY 121 SerThrValThrThrLeuValGlyGluMetAlaValAlaAlaAspLysArgValAlaVal 140
Db 291492 AGCACCGTGACACCGCTGTCGGCGAGATGGCGCGAGCTGCCGCGAAGCGTGTGCTGTG 291433
QY 141 GlyGlyAsnLeuGlyThrProAlaLeuAspLeuLeuAlaAspIleGluLeuTyrVal 160
Db 291432 GCGCGCAACCTGGCGCACCGCGCGTGGACCTGCTCAGCGATGATCGAGTGTGACGTG 291373
QY 161 LeuGluLeuSerSerPheGlnLeuGluThrCysAspArgLeuAsnAlaGluValAlaIleThr 180
Db 291372 ATGGAATTGCCAGTTTTCAGCTCGAAACCACTGATCAGCTCAATGCCGAAGTGGCGACC 291313
QY 181 ValLeuAsnValSerGluAspHisMetAspArgTyrAspGlyMetAlaAspTyrHisLeu 200
Db 291312 GTGCTGAACATCAGCAAGACCATATGGACCGTTATTCGCGTCTGCGCGCTATCACCTC 291253
QY 201 AlaLysHisArgIlePheArgGlyAlaArgGlnValValValValValValValValVal 220
Db 291252 GCCAAGCACCGTATTTTCGTCGGCGCGCGCAAGTAGTGTCAATCGGAGGAGCGCGCTG 291193
QY 221 ThrArgProLeuIleAlaAspThrValProCysTyrTrpSerPheGlyLeuAsnLysProAsp 240
Db 291192 TCCCGTCGTTGATCGCGGAGCGCTGCTGCTGTCGACGTTCCGCTCACACAGCTGAC 291133
QY 241 PheLysAlaPheGlyLeuIleGluGluAspGlyGlnLysTrpLeuAlaPheGlnPheAsp 260
```

```
Db 291132 TTCCACGGCTTGGTTCGCTGAAGAGAAGCGGAAAAATTCCTGTCATTCAGCTCGAG 291073
QY 261 LysLeuLeuProValGlyGluLeuLysIleArgGlyAlaHisAsnTyrSerAsnAlaLeu 280
Db 291072 AACCTGATCCGCTGAGCGAGCTGAAGTTGCGCGCGCAACAATCAGGTACGACCTG 291013
QY 281 AlaAlaLeuAlaLeuGlyHisAlaValGlyLeuProPheAspAlaMetLeuGlyAlaLeu 300
Db 291012 GCAGCCCTGGCGTGGGCGATCCGCTGGCGCTGCCGTTTCGACGCCATGCTTCCAGCCTG 290953
QY 301 LysAlaPheSerGlyLeuAlaHisArgCysGlnTrpValArgGluArgGlnGlyValSer 320
Db 290952 CGTGAAGTTCACCGGCTTTCGACCGCTGCCAATGCTGCTGAGCGCAATGCGCGTGGAC 290893
QY 321 TyrTyrAspSerLysAlaThrAsnValGlyAlaAlaLeuAlaIleGluGlyLeu 340
Db 290892 TACTACACGATTCCCAAGGCCACCAACGTTGGTGTGCTGCGCTATCGAAGGCCCTG 290833
QY 341 GlyAlaAspIleAspGlyLysLeuValLeuAlaGlyGlyAspGlyLysGlyAlaAsp 360
Db 290832 GGCTCGATATCGCGGCAAGCTGTTGATGCGCGGTGGCATGGCAAGGTGCGGAC 290773
QY 361 PheHisAspLeuArgGluProValAlaArgPheCysArgAlaValValLeuLeuGlyArg 380
Db 290772 TTACAGCGGCTGCTGCCCGCTGGCTAAATACTGCGCGCTGCTGCTGCTGCTGCTG 290713
QY 381 AspAlaGlyLeuIleAlaGlnAlaLeuGlyAsnAlaValProLeuValArgValAlaThr 400
Db 290712 GATGCGAGCTGATTCGCCAGCGCTGGGTGATGCGCTGCGCTGATTCGTGTCGACAG 290653
QY 401 LeuAspGluAlaValArgGlnAlaGluLeuAlaArgGluGlyAspAlaValLeuLeu 420
Db 290652 GTTCAGCGCGCTCCCGAGCGCGAGCTGGCTCAGCGCGCGATGCCGTGTTGCTT 290593
QY 421 SerProAlaCysAlaSerLeuAspMetPheLysAsnPheGluGluArgGlyArgLeuPhe 440
Db 290592 TCCTCTGCTATGCGCCAGCTGTCAGATGTTCAAGATTATGAGAGCGCGGACGTGTGTC 290533
QY 441 AlaLysAlaValGluGluLeuAla 448
Db 290532 GCGCAGCGAGTGGAGTGTGTGCA 290509

RESULT 5
LOCUS AE016960
DEFINITION Cxiella burnetii strain RSA 493, section 1 of 7 of the complete genome.
ACCESSION AE016960 AE016828
VERSION AE016960.1 GI:29540626
KEYWORDS Cxiella burnetii RSA 493
SOURCE Cxiella burnetii
ORGANISM Bacteria; Proteobacteria; Gammaproteobacteria; Legionellales; Cxiellaceae; Cxiella.
REFERENCE 1 (bases 1 to 301727)
AUTHORS Seshadri,R., Paulsen,I.T., Eisen,J.A., Read,T.D., Nelson,K.E., Nelson,W.C., Ward,N.L., Tettelin,H., Daviden,T.M., Beanan,M.J., DeBoy,R.T., Daugherty,S.C., Brinkac,L.M., Madupu,R., Dodson,R.J., Khouri,H.M., Lee,K.H., Carty,H.A., Scanlan,D., Heinen,R.A., Thompson,H.A., Samuel,J.E., Fraser,C.M. and Heidelberg,J.F.
TITLE Complete genome sequence of the Q-fever pathogen Cxiella burnetii
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 100 (9), 5455-5460 (2003)
PUBMED 12704232
REFERENCE 2 (bases 1 to 301727)
AUTHORS Seshadri,R., Paulsen,I.T., Eisen,J.A., Read,T.D., Nelson,K.E., Nelson,W.C., Ward,N.L., Tettelin,H., Daviden,T.M., Beanan,M.J., DeBoy,R.T., Daugherty,S.C., Brinkac,L.M., Madupu,R., Dodson,R.J., Khouri,H.M., Lee,K.H., Carty,H.A., Scanlan,D., Heinen,R.A., Thompson,H.A., Samuel,J.E., Fraser,C.M. and Heidelberg,J.F.
TITLE Direct Submission
JOURNAL Submitted (04-APR-2003) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA
```

FEATURES	Source	Location/Qualifiers
gene		<pre> 1. .301727 /organism="Coxiella burnetii RSA 493" /mol_type="genomic DNA" /strain="RSA 493" /db_xref="taxon:227377" 140. .1495 /gene="dnaA" /locus_tag="CBU0001" 140. .1495 /gene="dnaA" /locus_tag="CBU0001" </pre>
		<pre> /note="identified by match to PFAM protein family HMM pf03299" /codon_start=1 /transl_table=11 /product="chromosomal replication initiator protein DnaA" /protein_id="AA089571.1" /db_xref="GI:29540627" </pre>
CDS		<pre> /translation="MSLPTSLMDKCLGYLRDIPPOQYNTWIRPLHATESKONGLLLL APNRFLVDMINERFNIRITELDELSDTPPQIRLQIGRSTEMPTKNSHPSHRKAAA PPAGTITSHQTANINSEFTFDSFVEGKSNQLARAAATQVAENPGAYNPLFYGGVGL GKTHLHVAQNAILKDKSKVYLHDSREVDAMIKALQHNAMNEKRFYRSLNALLI DDIOFEACKORSOEFEFTFNALLDQGOQIILTCDRYPKEINLEERLQSRFGWGLTV ATEPELETRVAILMSKAEQLKVHLPEHVAFFIAKHQISNVRELEGALKRVIANAHFT GOSITVTEALKOLLTQARLIITENIQTKVAEYKIKKVADLLAKRNRSVARPRQ MAMALAKETNHSLEPEIGDAFGGRDHTTVHACRVKVELLATSILEDYKNLMRLS G" 1639. .2748 /gene="dnaA" /locus_tag="CBU0002" 1639. .2748 /gene="dnaA" /locus_tag="CBU0002" </pre>
		<pre> /note="identified by match to PFAM protein family HMM pf02767" /codon_start=1 /transl_table=11 /product="DNA polymerase III, beta subunit" /protein_id="AA089572.1" /db_xref="GI:29540628" </pre>
gene		<pre> /translation="MKNLIRETLKPLQVIGVVERKQTLPILSNVLISLETDOLSI TGTDLLEVELIGTKLDKNSTESSRLTLGRKLMDCRALPDNAPIELYRDREKILRS GSRFMLSLETSDEDPAYERRKSELELALPOSTFNHLLYRTHFMAAQDVRYYLNGLL LETYPNKRATADGHLAANTLLQVNTNHLQIIPRKGIIEILLRLEDEALATC RIGNNHGVSTNDEFTTSLKIEGFPDCERVIPKGGDKQFVIDRDLKQALSRTAILC NEKFGVRFELROGLRILATNPEOAAEEINIDYTCENLDIGNFVGVLDILNVVN SGNIRLFTSDSSVLINEVENPDSAFVMPRL" 2723. .3826 /gene="recF" /locus_tag="CBU0003" 2723. .3826 /gene="recF" /locus_tag="CBU0003" </pre>
		<pre> /note="identified by match to PFAM protein family HMM pf02463" /codon_start=1 /transl_table=11 /product="recF protein" /protein_id="AA089573.1" /db_xref="GI:29540629" </pre>
gene		<pre> /translation="MWLCICAYSRMPYIGSLKVNQFRNLADVDTIRHSQFNFFGQNG AGKTSIGESIYLSVRSFRTHLPQRLIQDNTDRFLIFITLYNGTQFIPLGVERDCHG DRCRLINGETASSSLAAKRLPLCLSLMSHRELLDGPVRVRQFLDWLMPHFVEPSFSS IWORLSLKORNAALKAKPLPGETIHWDKMLVEDGERLHOLRONVYTERKPLETQML QOFLPVLIGHYFRGSEKYSIMEQLQINLKQDLRGYTAGQQRADFRLTIGDLPA QDLISQGOQKLYTALHFAQGLLKENTGISPIYLIDDLPAELDANKRDCVIDLVNCL ESQVIFSGIDPNELRLPHPSLFLFHVKHGKVAAL" 4128. .6554 /gene="gyrB" /locus_tag="CBU0004" 4128. .6554 /gene="gyrB" </pre>
		<pre> /note="identified by match to PFAM protein family HMM pf01402" /codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="AA089577.1" /db_xref="GI:29540633" </pre>
gene		<pre> /translation="MKKEYDFSRAKNPNYAKGLKQVTIRLDAETINFYKRLAEDSNI PYOMLINLYRECAKHKHKLNISWEK" 8716. .8904 /locus_tag="CBU0008" 8716. .8904 </pre>
		<pre> /note="identified by match to PFAM protein family HMM pf02518" /codon_start=1 /product="DNA gyrase, B subunit" /protein_id="AA089574.1" /db_xref="GI:29540630" </pre>
gene		<pre> /translation="MSVSAPOKYDSSTIKVLKGLDAVRKRPMYIGDITDDGSLHHMV FEVDNSIDEALAGRCFSHIWTIHADGSVSRDNGRGIPDTIHPEGRSAAEVINTVL HAGCKEDANAYKVGSLHGVGSVYNALSEELYMTIRONGVYVYQTYKMGVQAPLAV TKSKXTCTEIRFRPSNALTFRNTERHYDILAKRLRELAFNLGVRIELTDERSGRDI FEYEGIRAFVEHLNKKPTIHSDFHSTENKNDIVIVAMOWNDFQETIFCFNTNI PQDGGTHLAGFRALTRTLANNYIEGEALAKSKVTISGDDTREGTLTVLSVKVPDPK FSSQTKDKLVSSEVPVVESTAKOHDYLLLENPKQARIAMKVIDAARAREAKARK DTRRKNVLDLGGPLSKLADCEKDPALCEFLVEGDSAGGSAQARRRYQAILPLK GKILNVEKARFDKMLSSAEGTLITALCGGIGKSEYNPKLRYHRIILMTDADVDGSH IRTLLLFFYQMPLELIERGHYIIAQPLPYKVKKSQEQYVDDALLALLQLALDG AQFHTPOAPAIQGSALENLAAHYNTAMRSINRLGKRYPLYILKRLNFPPLTVADLS BEEKVIANTQBLEQNLNPEKNSLYQTAIKDAERELYLPMTYIINHGIAQOHLFH QEFQGCADYRHLVDLYNKISGLIEPGAYIVRNDKRTVTHFEALWETIEGKGLNI QRYKGLGEMNPDQLWETMTDRESRLLSVSIEDAVMADEIFTTLMGDQVEPRRLFIEQ NALAVENLDV" </pre>
		<pre> complement(6528. .6722) /locus_tag="CBU0005" complement(6528. .6722) /locus_tag="CBU0005" /note="identified by Glimmer2; putative" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="AA089575.1" /db_xref="GI:29540631" </pre>
gene		<pre> /translation="MAHYDAVFKGWVHFGLLVENLHFQSHPLVGNDMGSPFLEGSDS GSFYFYLELGLLNVQLHR" 6759. .7778 /locus_tag="CBU0006" 6759. .7778 </pre>
		<pre> /locus_tag="CBU0006" /locus_tag="CBU0006" /note="similar to PIR:A42937, GB:X66899, GB:X72990, GB:X72991, GB:X72992, GB:X72993, GB:X72994, GB:X72995, GB:X72996, GB:X72997, GB:X72998, GB:X72999, GB:X73000, GB:X73001, GB:X73002, GB:X73003, GB:X73004, SP:001844, PID:1572086, PID:1666068, PID:31280, and PID:825654; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="IS1111A, transposase" /protein_id="AA089576.1" /db_xref="GI:29540632" </pre>
gene		<pre> /translation="MKDIKILGVDIADKDVFCGLDEMGKVIYTRRVKRAQYVSTVAS LKVCVYMEACGGANHWYRTFMGIGIPTQLISPOHVPYVSKNDRNDQAIASAAAS RASMRVQGTVEQDDYQALLKIDRLVKSFTALINERGLLOVGLTMARCAKRFYE ELPILASEAVGLTPRMKRVNLNCLYTELLNDEALIGDYEEELKAVAKANEDCQRVSI PGVGYLTALSVYASGDIHOFHSRQLSAFGLVPRQHSNGKNKEVLGISKRGVNLRL TLLIHGARALLRHVKNKRTDKLSLWKALIERGRNMNRACVALANKNAPIIALLTROET YRCGA" </pre>
		<pre> complement(8186. .8398) /locus_tag="CBU0007" complement(8186. .8398) /locus_tag="CBU0007" /note="identified by match to PFAM protein family HMM pf01402" /codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="AA089577.1" /db_xref="GI:29540633" </pre>

```

/locus_tag="CBU0008"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA089578.1"
/db_xref="GI:29540634"
/translation="MSNFQKQKRAVSDSPALSNFFISVIFKSVTSIPVLSNWRRTN
PPKPPKLKLTGGRATFVS"
9560..10993
/gene="dacB"
/locus_tag="CBU0009"
9560..10993
/gene="dacB"
/locus_tag="CBU0009"
/notes="identified by match to PFAM protein family HMM
PF00905"
/codon_start=1
/transl_table=11
/product="D-alanyl-D-alanine
carboxypeptidase/D-alanyl-D-alanine-
endopeptidase"
/protein_id="AA089579.1"
/db_xref="GI:29540635"
/translation="MKPPYLAILLSFCPLFAVEKQNPOTNINILSIKSKANIG
FIVSDVKTRVIFERSQELFSPASTOKLFTVAVALYYLGSDFYFTALLTNGAIRGQ
TLQGNLTLPFGSDPELTEDLNRLIEKLKELGHRISGHVYIDNTAYNDVYPPGWWL

```

## Alignment Scores:

```

Pred. No.: 1 58e-55 Length: 301727
Score: 1023.00 Matches: 210
Percent Similarity: 64.47% Conservativeness: 84
Best Local Similarity: 46.05% Mismatches: 140
Query Match: 45.27% Indels: 22
DB: 1 Gaps: 3

```

## US-09-701-229-2 (1-448) x AE016960 (1-301727)

```

Qy 2 SerLeuIleAlaSerAspHisPheArgIleValValGlyLeuGlyLysSerGlyMetSer 21
    ::::::::::::::::::::
Db 118030 ACGTATTAGTAGTACGAATCATTAACGGTTATCGTTGGTTAGTAAACGGGATTATCT 118089
Qy 22 LeuValArgTyrLeuAlaArgGlyLeuProPheAlaValValAspThrArgGluAsn 41
    ::::::::::::::::::::
Db 118090 TGTGCGGAGTTTTTAGCGAAGAACACCGCGTTTGGCGGTGATGGACAGTCGTGAAGAA 118149
Qy 42 ProProGluLeuAlaThrLeuArgAlaGlnTyrProGlnValGluValArgCysGlyGlu 61
    ::::::::::::::::::::
Db 118150 CCCCCGAGTGGGAAATTTTATCAAACTTACCTCGCGTCGAATTGATCCGCGGCCAA 118209
Qy 62 LeuAspAlaGluPheLeuCysSerAlaArgGluLeuTyrValSerProGlyLeuSerLeu 81
    ::::::::::::::::::::
Db 118210 TTTTCGGAATAATTATTGAACGAAGCGCAAGAGATTATTAAAGTCCTGCGGTTTCGCTG 118269
Qy 82 ArgThrProAlaLeuValGlnAlaAlaLysGlyValArgIleSerGlyAspIleAsp 101
    ::::::::::::::::::::
Db 118270 CAGAGCGGTAAATTCGCAACGAAGCAGCGCAAGCAAAATTCATATTATGGCGATATTGAA 118329
Qy 102 LeuPheAlaArgGluAlaLysAlaProIleValAlaIleThrGlySerAsnAlaLysSer 121
    ::::::::::::::::::::
Db 118330 TTATTCGACGCAATGTGATAAACCAATTATTGCTATTACCGGTCCTCAATGGCAAAACC 118389
Qy 122 ThrValThrThrLeuValGlyGluMetAlaValAlaAspLysArgValAlaValGly 141
    ::::::::::::::::::::
Db 118390 ACCGTGACGCGGTAGTGGGTAATGATGAAAGCCCGCGCAACAGTTTCAGTTTGC 118449
Qy 142 GlyAsnLeuGlyThrProAlaLeuAspLeuAlaAspIleGluLeuTyrValLeu 161
    ::::::::::::::::::::
Db 118450 GGTAATATGGCGGCGGCTTTAGACAAATATACCTCAGCGCGGATTATACGTACTG 118509
Qy 162 GluLeuSerSerPheGlnLeuGluThrCysAspArgLeuAsnAlaGluValAlaThrVal 181
    ::::::::::::::::::::
Db 118510 GAATATCGAGCTTCCAACTTGAAACGACATTTTCTTTTACGATCTCAAGCGGCAACATA 118569

```

```

Qy 182 LeuAsnValSerGluAspHisMetAspArgTyrAspGlyMetAlaAspTyrHisLeuAla 201
    ::::::::::::::::::::
Db 118570 TTAATATTAGCGAAGACCATATGAATCGTATCGGACTTATCAAGATTATCTCCGAGCT 118629
Qy 202 LysHisArgIlePheArgGlyAlaArgGlnValValValAsnArgAlaAspAlaLeuThr 221
    ::::::::::::::::::::
Db 118630 AAACAACGGCATTTATCTAGTTGTTTATTCCTCATTTGTAAC----- 118671
Qy 222 ArgProLeuIleAlaAspThrValProCysTrp----- 232
Db 118672 -----CGGACGACCAAGCAAAATTTGGGCCCATCTTCGCTTTAATAAAAAACCT 118719
Qy 233 ---SerPheGlyLeuAsnLysProAspPheLysAlaPheGlyLeuIleGluAspGly 251
    ::::::::::::::::::::
Db 118720 TTGCTCTTCGGGTTAAAT-----AATGCTGCTGATTTTCGTTGGCAGAACATAATCAA 118773
Qy 252 GlnLysTrpLeuAlaPheGlnPheAspLysLeuProValGlyLeuLysIleArg 271
    ::::::::::::::::::::
Db 118774 AAAACGCTCTATTGCTCATCATCGGTAAGATTTTAATGCCCATCCAAGAAATTTAAATTAAC 118833
Qy 272 GlyAlaHisAsnTyrSerAsnAlaLeuAlaLeuAlaLeuGlyHisAlaValGlyLeu 291
    ::::::::::::::::::::
Db 118834 GCCCGTCATCATCTCCAAATGTCACCTGCTGCTATGACATTGGGACAGCAGCTAAATA 118893
Qy 292 ProPheAspAlaMetLeuGlyAlaLeuLysAlaPheSerGlyLeuAlaHisArgCysGln 311
    ::::::::::::::::::::
Db 118894 CCCATAGAAAATATGCTTTCATGATTATACGTGATTTCGCGTATCCGCCACCGCTGTCAA 118953
Qy 312 TrpValArgGluArgGlnGlyValSerTyrTyrAspAspSerLysAlaThrAsnValGly 331
    ::::::::::::::::::::
Db 118954 TGGGTACGCAAGTATAGGAAATTTGATTATTAATGATTCAAAAGGCACAAACGTTGGC 119013
Qy 332 AlaAlaLeuAlaAlaIleGlyLeuGlyAlaAspIleAspGlyLysLeuValLeuLeu 351
    ::::::::::::::::::::
Db 119014 GCTACCGCGCTGCCATTGAAGCTTAGGCAAGCGCAAAAGGCAACTTAATCCTATC 119073
Qy 352 AlaGlyCysAspGlyLysGlyAlaAspPheHisAspLeuArgGluProValAlaAlaArgPhe 371
    ::::::::::::::::::::
Db 119074 GCTGGGGACAAGCAAGGCGCGGATTTTCTCCATTAAAGACGTCGTTAAACGTTAC 119133
Qy 372 CysArgAlaValValLeuLeuGlyArgAspAlaGlyLeuIleAlaGlnAlaLeuGlyAsn 391
    ::::::::::::::::::::
Db 119134 GTTAACAAGTGATTTTAATCGCGAGGATGCTCCGCTGTGTAAGAAAACCTTAAGAA 119193
Qy 392 AlaValProLeuValArgValAlaThrLeuAspGluAlaValArgGlnAlaAlaGluLeu 411
    ::::::::::::::::::::
Db 119194 ATAACCGTTATTAAACACGCTGATTCAATGAACGACCGCTCAACGATCCATCAAGCC 119253
Qy 412 AlaArgGluGlyAspAlaValLeuLeuSerProAlaCysAlaSerLeuAspMetPheLys 431
    ::::::::::::::::::::
Db 119254 GCCAAAGCCGCGATATTGTTTATTATCCCGCTTGCAGAGTTTGTATGATGTTTACC 119313
Qy 432 AsnPheGluGluArgGlyArgLeuPheAlaLysAlaValGluLeu 447
    ::::::::::::::::::::
Db 119314 AACTACGAACACCGCGCGATGTTTTTACAGAAACGCTGGAAGCGCTT 119361

```

## RESULT 6

```

AL627265 251050 bp DNA linear BCT 06-JUN-2002
LOCUS Salmonella enterica serovar Typhi (Salmonella typhi) strain Crl8,
DEFINITION complete chromosome, segment 1/20.
ACCESSION AL627265 AL513382
VERSION AL627265.1 GI:16501283
KEYWORDS
SOURCE
ORGANISM
Salmonella enterica subsp. enterica serovar Typhi
Salmonella enterica subsp. enterica serovar Typhi
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Salmonella.
1 (bases 1 to 251050)
Parkhill, J., Dougan, G., James, K. D., Thomson, N. R., Pickard, D.,
Wain, J., Churcher, C., Mungall, K. L., Bentley, S. D., Holden, M. T. G.,
Sebahia, M., Baker, S., Basham, D., Brooks, K., Chillingworth, T.,
Conerton, P., Cronin, A., Davis, P., Davies, R. M., Dowd, L., White, N.,

```



Farrar, J., Feltwell, T., Hamlin, N., Haque, A., Hien, T. T., Holroyd, S., Jagels, K., Krogh, A., Larsen, R. S., Leather, S., Moule, S., O'Gaora, P., Parry, C., Quail, M., Rutherford, K., Simmonds, M., Skelton, J., Stevens, K., Whitehead, S. and Barrell, B. G.  
Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar Typhi C718  
Nature 413 (6858), 848-852 (2001)  
21534947  
11677608  
2 (bases 1 to 251050)  
Parkhill, J.  
Direct Submission  
Submitted (25-Oct-2001) Submitted on behalf of the Salmonella sequencing team, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK  
E-mail: parkhill@sanger.ac.uk  
Notes:  
Details of *S. typhi* sequencing at the Sanger Centre are available on the World Wide Web.  
(URL, [http://www.sanger.ac.uk/Projects/S\\_typhi/](http://www.sanger.ac.uk/Projects/S_typhi/)).

## FEATURES

## source

1. 251050  
/organism="Salmonella enterica subsp. enterica serovar Typhi"  
/mol\_type="genomic DNA"  
/strain="C718"  
/db\_xref="taxon:90370"  
190. 255  
/gene="STY0001"  
/note="synonym: thrL"  
190. 255  
/gene="STY0001"  
/note="Orthologue of *E. coli* thrL (LPT\_ECOLI); Fasta hit to LPT\_ECOLI (21 aa), 86% identity in 21 aa overlap"

## gene

## CDS

/codon\_start=1  
/product="thr operon leader peptide"  
/protein\_id="CAD01154.1"  
/db\_xref="GI:16501284"  
/db\_xref="SPTREMBL:Q8XG12"  
/translation="MNRISTTITITITITGGAG"  
337. 2799  
/gene="STY0002"  
/note="synonym: thrA"  
337. 2799  
/gene="STY0002"  
/note="Orthologue of *E. coli* thrA (AK1H\_ECOLI); Fasta hit to AK1H\_ECOLI (820 aa), 94% identity in 820 aa overlap"

## gene

## CDS

/codon\_start=1  
/product="aspartokinase I/homoserine dehydrogenase I"  
/protein\_id="CAD01155.1"  
/db\_xref="GI:16501285"  
/db\_xref="SPTREMBL:Q8Z97"  
/translation="NRVLKFGTGSNAERELRVADILLESROGQVATVLSAPAKIT NHLVAMIEKTIQQDALPNISDAERIFSDLLAGLASAQPGFLARKKVMQDEFAQIK VLRGISLGGQCPDSINALICRGKMSIATMAGLEAREHVRVTIDVPEKLVAGHY LESTVDIAESTRIAASQIPADHMTLMAGFTAGNGELVLRNGSDYSAAVLAAQY RADCCIEIWTVDVGYTCDPQVPDARLLKMSYQEAELSFGAKVLHPRTITPAQF QTPCLIKTGNPQAPGTLTGASDDNLPVKGISNLNNMAMSVSGPGMKMGIMGAAR VFAMSRAGISVLLVITQSSSEYISFCVPSQDCARARMQDEFEYLEKGLLEPLAV TERLAIISVGGMHTLRGISAKFFAALARANINIVATAQSSERSISVYVNNDDATT GVRVTHQMFNDQVIEVFVIGVGGVGGALLEQKQQTWLKNKHIDURVCGVANSKA LITNVHGLNWNQAEALQANAPFNGLRIILVKEVHLLNPFVVDCTSSQAVOAYAD FLREGFVVTNPKANTSMFYHQLEFAAQSRKFLYDYNVAGLFPVIEINQLNLN AGDELFKFGSLGSLSEIFGKLEEGMSLQATALAREMGYTEPDDRLSGMDVARK LLILARETGLELSDIVIEPVLDFEDASGDVTFAMHLQDLDFAFAVAKARDEC KVLRVGNEEDGVCVKIAEWDGNDPLFKVNGENALAFYSHYQPLPLVLRGVGAG NDVTAAGVADLRLTLRWKLG"  
337. 1101  
/gene="STY0002"  
/note="Pfam match to entry PF00696 aakinease, Amino acid kinase family, score 207.70, E-value 1.7e-58"

## misc\_feature

## misc\_feature

## misc\_feature

## misc\_feature

## gene

## CDS

343. 369  
/gene="STY0002"  
/note="PS00324 Aspartokinase signature"  
1819. 2781  
/gene="STY0002"  
/note="Pfam match to entry PF00742 Homoserine\_dh, Homoserine dehydrogenase, score 535.60, E-value 3.4e-157"  
2314. 2382  
/gene="STY0002"  
/note="PS01042 Homoserine dehydrogenase signature"  
2801. 3730  
/gene="STY0003"  
/note="synonym: thrB"  
2801. 3730  
/gene="STY0003"  
/note="Orthologue of *E. coli* thrB (KHSE\_ECOLI); Fasta hit to KHSE\_ECOLI (310 aa); 94% identity in 308 aa overlap"

## CDS

/codon\_start=1  
/transl\_table=11  
/product="homoserine kinase"  
/protein\_id="CAD01156.1"  
/db\_xref="GI:16501286"  
/db\_xref="SWISS-PROT:Q8XGP5"  
/translation="MVKYAPASSANMSVGFVDVGAATVPDGTLLGDVSVVEADHF RLNLGREFADKLPPPEPNIYQCWERECQALGKTIPTVAMLEKNMPTGSLGSSACS VVAALVAMNEHCQKPLNDTRLIALMGELEGRISGSIHYDNDVAPCFLGQMILIEGNI ISOQVPGFDEMLWVLAIPGIVKIVTAERAILPAQYRRDCCIAHGRHLAGFTHACYSRQ POLAALMKDVIAEYPRARLLPGFSQARQAVSEIGALASIGSGSTFLFALCDKPEFA ORVADWLKSHVQLNQEGFVHCRLDTAGRVVG"  
3005. 3196  
/gene="STY0003"  
/note="Pfam match to entry PF00288 GHMP\_Kinases, GHMP kinases putative ATP-binding proteins, score 79.20, E-value 5.7e-21"  
3068. 3103  
/gene="STY0003"  
/note="PS00627 GHMP kinases putative ATP-binding domain"  
3734. 5020  
/gene="STY0004"  
/note="synonym: thrC"  
3734. 5020  
/gene="STY0004"  
/note="Orthologue of *E. coli* thrC (THRC\_ECOLI); Fasta hit to THRC\_ECOLI (428 aa), 93% identity in 428 aa overlap"

## misc\_feature

## misc\_feature

## gene

## CDS

/codon\_start=1  
/product="threonine synthase"  
/protein\_id="CAD01157.1"  
/db\_xref="GI:16501287"  
/db\_xref="SPTREMBL:Q8Z9R6"  
/translation="MKLYNLKDHNEQVSFAQVQTGLGKQOGLFFPHDLPEFSLTEID EMLNQDFVRSKILISAFIDGDEIPQIILEERVRAAFAPAPVAQVESDGLGELFHGP TLAQDFGGRFMAQMLTHISGDKPVTILTSGDTGAVAHAFYGLENNRVVILYPRG KISLPQELKCTLGSIETVAIDGDFDCAQALVKQADDEELKTKLAGNSANSINISR LLAQCIYFEAVAQIPQAGARNQLVISVPSGNGDLTAGLLAKSLGLPKVRTIAATNIN LTAQPRFLHDGKWAQPTATLSNADWVSQPNWPRVEELFRKKIWRULFELZGYAAVDDT TQQTQRELKAGYIISEPHAAYRALRDQNPGEYGLFLGTAHPAKFKESVESILGE TLALPEALAEADRLPLLSHLHPADFAALRKLMMTRQ"  
3962. 4867  
/gene="STY0004"  
/note="Pfam match to entry PF00291 PALP, Pyridoxal-phosphate dependent enzyme, score 104.50, E-value 2e-27"  
4022. 4066  
/gene="STY0004"  
/note="PS00165 Serine/threonine dehydratases pyridoxal-phosphate attachment site"  
complement(5114. 5887)  
/gene="STY0005"  
/note="synonym: yaaA"  
complement(5114. 5887)  
/gene="STY0005"  
/note="Orthologue of *E. coli* yaaA (YAAA\_ECOLI); Fasta hit

## misc\_feature

## misc\_feature

## gene

## CDS



to YAAA\_ECOLI (258 aa), 86% identity in 257 aa overlap"

```
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="CAD01158.1"
/db_xref="GI:16501288"
/db_xref="SPIREMBL:Q8Z9R5"
/transl_table="MELIISPAKTLDYQSLATRYTQPELLDHSQQILQOARQLSP
QISLMISKLADLNATRFDMQPHFTPDNARQAILAFKGDVYTGQAETFFNDADF
FAOHLRLMGLVLEIRLEMLPRKDYOFWGDITDKLEALE
AQDGVVNLASREYFKSKPKLDALIKPFLDEKNGFKVSVFYAKKARGLMSRE
LIENRLKPOLAFADREGYFDEETSTQDELVEKRYEQ"
complement(5966..7396)
/gene="STY0006"
/note="synonym: yaaq"
complement(5966..7396)
/gene="STY0006"
```

gene

CDS

```
/note="Similar to Bacillus subtilis amino acid carrier
protein alst ALST SW:ALST_BACSU (Q45068; P40743) fasta
scores: E(): 0, 45.3% id in 477 aa, and to Alteromonas
haloplanktis na(+)-linked D-alanine glycine permease DAGA
SW:DAGA_ALTHA (P30144) fasta scores: E(): 0, 36.5% id in
315 aa"
```

Orthologue of E. coli yaaJ (YAAJ\_ECOLI); Fasta hit to  
YAAJ\_ECOLI (476 aa), 76% identity in 476 aa overlap"

```
/codon_start=1
/transl_table=1
/product="putative amino-acid transport protein"
/protein_id="CAD01159.1"
/db_xref="GI:16501289"
/db_xref="SPIREMBL:Q8Z9R4"
```

```
/transl_table="MPEFFSFINELMGSMVYLLGAGCWFTWRTGFIQFRYIQFS
RSKLSFSPQGGTSLQCTSLAARGSGNLGVALAIAAGGPGAVFWMYSAIIG
MATSFACSLAQLYKEDPTQGFRRGPGAWMARGLMWVFAFLVAYGLIENS
VOANAVSRHLAFNIPPLISGIALAFCALLIIRGKIGVARMLOWILPIALILWAG
SVFTCLWHIQMGPVIASIVKSAFGWOERAGACGYTLQATISGFORGEENRAG
STPNAAAATSYPPHPVAQGIQVOMIGVFSDTIICITASMIILLAGNHASHSTEGIQ
LQHAMVLTGEMGAEVALIVILFAFSSIVANTYIAENNLFFLRHNKAIAWILRIA
TLGHVIAVTLISPLIQLADIMINMACMAITNLTAILLSPVYVTLAGDYLRLKLGVR
```

#### Alignment Scores:

Pred. No.:	1.83e-53	Length:	251050
Score:	990.00	Matches:	233
Percent Similarity:	64.18%	Conservative:	59
Best Local Similarity:	51.21%	Mismatches:	139
Query Match:	43.81%	Indels:	24
DB:	1	Gaps:	10

US-09-701-229-2 (1-448) x AL627265 (1-251050)

Qy	2	SerLeuIleAlaSerAspHisPheArg	-----lleValValGlyLeuGlyLysSer	18
Db	148805	ACGTTAATCATGGCTGATTACACGAGCAAAAACGTCGTCATATATCGTCTGGGCTTAACC	148864	
Qy	19	GlyMetSerLeuValArgTyrLeuAlaArgArgGlyLeuProPheAlaValAlaValAspThr	38	
Db	148865	GGACTCTCTCGTGGAGCTTTTCTCGCCCGCGCGGTGACGCCGCGGTGATGGATACT	148924	
Qy	39	ArgGluAsnProProGluLeuAlaThrLeuArgAlaGlnTyrProGln	---ValGluVal	57
Db	148925	CGCTGACGCCGCGGCTCTGGATAAGCTG	-----CCGCAAGAGGTGGAGCGT	148972
Qy	58	ArgCysGlyGluLeuAspAlaGluPheLeuCysSerAlaArgGluLeuTyrValSerPro	77	
Db	148973	CACGTTGGCGCTGACGAGCAGTGCTCTTACGGCGGATTAATCGTCGCCAGCCCT	149032	
Qy	78	GlyLeuSerLeuArgThrProAlaLeuValGlnAlaAlaLysGlyValArgIleSer	97	
Db	149033	GGTATTGCTCGCGCATCCCTCTTAGCGCTCCGCCAGCGCTGCGAGTGGAATATGTC	149092	
Qy	98	GlyAspIleAspLeuPheAlaArgGluAlaLysAlaProIleValAlaIleThrGlySer	117	
Db	149093	GGCGATATCGAATCTGTTTGGCGGAGCAAGCGCGGATTTGTGGCCATCACCGGCTCG	149152	

Qy	118	AsnAlaLysSerThrValThrThrLeuValGlyGluMetAlaValAlaAlaAspLysArg	137
Db	149153	AACGGCAAAAGCACCGTGACCACTTAGTGGCGGAGATGGGAAGCGCGGCGTCAAT	149212
Qy	138	ValAlaValGlyGlyAsnLeuGlyThrProAlaLeuAspLeuLeuAlaAspIleGlu	157
Db	149213	GTCGGCGTTGGCGCAATATCGGTCTTCGCGCGTGTGCTGTGGATGCGGATCGCGAA	149272
Qy	158	LeuTyrValLeuGluLeuSerSerPheGlnLeuGluThrCysAspArgLeuAsnAlaGlu	177
Db	149273	TTGTACGTGCTGGAATATTCAGTTTCAATTTGAAATACCTCAAGTTTGCACGCGCG	149332
Qy	178	ValAlaThrValLeuAsnValSerGluAspHisMetAspArgTyrAsp	196
Db	149333	CGCGCAACGGTGCTCAACGTCACCTGAAGATCATATGACCGGTATCGGTTTGTGTCGA	149392
Qy	197	AspTyrHisLeuAlaLysHisArgIlePheArgGlyAlaArgGlnValValAlaAsnArg	216
Db	149393	CAGTACCGCGCGGCGGCAACTGCGGTCTACAGAGGCGGAAAGTGTGCTAGTAGTGC	149452
Qy	217	AlaAspAlaLeuThrArgProLeuIleAlaAspThrValProCysTyrSerPheGlyLeu	236
Db	149453	GATGACCGGTGACGATGCGGTACGCGGGGCCCATGAGCGCTGCTGAGCTTTGGCGTC	149512
Qy	237	AsnLysProAspPheLysAlaPheGlyLeuGluGluAspGlyGlnLysTyrLeuAla	256
Db	149513	AATATGGGTGATTATCAC-----CTTAATCGTCAGCAGGCGGAAACCTGCTACGA	149563
Qy	257	PheGlnPheAspLysLeuLeuProValGlyGluLeuLysIleArgGlyAlaHisAsnTyr	276
Db	149564	GTCAAAGGTGAGAAGGTGCTGAACGTGAAGAGATGAAGCTTCGGTTCAGCATAACTAT	149623
Qy	277	SerAsnAlaLeuAlaLeuAlaLeuGlyHisAlaValGlyLeuProPheAspAlaMet	296
Db	149624	ACCATCGGTAGCGCGCTGCGCTGCGGATGCCGTAGTGTGCTGCGCGCGCAGCAGT	149683
Qy	297	LeuGlyAlaLeuLysAlaPheSerGlyLeuAlaHisArgCysGlnTyrPvalArgGluArg	316
Db	149684	TTGAAGCGTTGACGACATTTACCGGTCTCGCGCATCGCTTCCAGCTGGCGTGGAGCAT	149743
Qy	317	GlnGlyValSerTyrTyrAspSerLysAlaThrAsnValGlyAlaAlaLeuAlaAla	336
Db	149744	AACGGCGTTCGTCGATCAACGACTCGAAAGCGACCAATGTGCGCAGTACGGAAGCGCG	149803
Qy	337	IleGluGlyLeuGlyAlaAspIleAspGlyLysLeuValLeuLeuAlaGlyGlyAspGly	356
Db	149804	CTAAACGGTTTG-----CATGTGACGGTACGTGCTGCTGCTGCGCGCGCAGCGT	149857
Qy	357	LysGlyAlaAspPheHisAspLeuArgGluProValAlaArgPheCysArgAla	374
Db	149858	AAGTCGCGACACTTT-----TCTCCGCTGGCGCTATCTGACCGCGCATCGT	149905
Qy	375	-----ValValLeuLeuGlyArgAspAlaGlyLeuIleAlaGlnAlaLeuGlyAsnAla	392
Db	149906	ATCGCGCTGATATGCTTTGGCGCGATGGCGCGCAGCTGTC---GCATCGCTCCGGAA	149962
Qy	393	ValProLeuValArgValAlaThrLeuAspGluAlaValArgGlnAlaAlaGluLeuAla	412
Db	149963	ATC---GCCCAACAGACTGAGACGATGGAAGCGGATGCGCTTTGCTGGCGCCCGCAGCTT	150019
Qy	413	ArgGluGlyAspAlaValLeuLeuSerProAlaCysAlaSerLeuAspMetPheLysAsn	432
Db	150020	CAGCGCGGTGATATGCTGCTGTCGCCCGCTGCGCCAGCTCGATCATCAGTTTAAAAAT	150079
Qy	433	PheGluGluArgGlyArgLeuPheAlaLysAlaValGluGluLeu	447
Db	150080	TTTGAGCAACGGGCGGATGCTTTACCGCTCTGCGGAAGGAGTTA	150124

RESULT 7

AE016834

LOCUS

DEFINITION Salmonella enterica subsp. enterica serovar Typhi Ty2, section 1 of

```

16 of the complete genome.
AE016834 AE014613
VERSION
AE016834.1 GI:29136167
SOURCE
ORGANISM
Salmonella enterica subsp. enterica serovar Typhi Ty2
Salmonella enterica subsp. enterica serovar Typhi Ty2
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Salmonella.
1 (bases 1 to 300169)
Deng, W., Liou, S.-R., Plunkett III, G., Mayhew, G.F., Rose, D.J.,
Burland, V., Kodoyianni, V., Schwartz, D.C. and Blattner, F.R.
Comparative Genomics of Salmonella enterica Serovar Typhi Strains
Ty2 and CT18
J. Bacteriol. 185 (7), 2330-2337 (2003)
2 (bases 1 to 300169)
Deng, W., Liou, S.-R., Plunkett, G. III, Mayhew, G.F., Rose, D.J.,
Burland, V., Kodoyianni, V., Schwartz, D.C. and Blattner, F.R.
Direct Submission
Submitted (25-SEP-2002) Laboratory of Genetics, University of
Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA
Location/Qualifiers
1. 300169
/organism="Salmonella enterica subsp. enterica serovar
Typhi Ty2"
/mol_type="genomic DNA"
/strain="Ty2"
/serovar="Typhi"
/sub_species="enterica"
/db_xref="taxon:209261"
190..255
/gene="thrL"
/locus_tag="t0001"
190..255
/gene="thrL"
/locus_tag="t0001"
/notes="corresponds to STY0001 from Accession AL513382:
Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="thr operon leader peptide"
/protein_id="AA067735.1"
/db_xref="GI:29136168"
/translation="MNRISTITITITITITGTGAG"
337..2799
/gene="thrA"
/locus_tag="t0002"
337..2799
/gene="thrA"
/locus_tag="t0002"
/notes="corresponds to STY0002 from Accession AL513382:
Salmonella typhi CT18; homoserine dehydrogenase I"
/codon_start=1
/transl_table=11
/product="aspartokinase I"
/protein_id="AA067736.1"
/db_xref="GI:29136169"
/translation="MRVLKFGTSVANAERFLRVADILESRSRQGVATVLSAPAKIT
NHLVAMIKTIGGDALPNISDAERIFSDLLAGLASAQPGLRLKMWVQEFQAQIK
HYLGISILGQCPDSINAALICRGKMSIATMAGLEAGRHRTVVDPEKILVAGHY
LFSTVDIAASTRAISOIPADHMTLMAGFTAGNEKGYLVLRNGSDYSRAVLAACL
RADCEIETWDGVYTCDPQVPDARLLKMSYQAMELSIFGARLVHPTITPIAQF
QIPCLIKTGNPACTLIGASDDNLPVKGISNLNNMFPVSGPMKGMIGMAAR
VFAAMRAGISVVLITQSSEYSISFCVQSDCARARMODEFYLEKLEGLLEPLAV
TERLAIISVVDGMDTLRGISAKFAALARANINIVAIAQSSERSISVWNDDATT
GVRYTHQMLFDQVIEVYIGVGVGALLLEOLKQOQTLKKNHIDLRVCGVANSKA
LITNVHGLNDNQALQAQAPNLRILRVKEVHLINPVIVDCTSSQAVADQYAD
FLREGFHVYTPNKANTSMDYHQLRFAAQRKFLIDYTVNGAGLPIENLQNLN
AGDELQKSGILSGSLSTIFGKLEGMSLSQATALAREMGTEPPDRDLDSGMQVARK
LILARETGRELELSDIVIEPVLDPDEFDAGDVTAFMAHLPLQDDAFARVAKARDEG
LVYRVGVNIEEDGVGRVIAEIVDGNLPLFKVKNGENALAFYSHYYQPLPLVLRGYGAG
NDVTAAGVFADLLRLTSNKLGV"

```

gene

CDS

2801..3730

/gene="thrB"

/locus\_tag="t0003"

2801..3730

/gene="thrB"

/locus\_tag="t0003"

/note="corresponds to STY0003 from Accession AL513382:

Salmonella typhi CT18"

/codon\_start=1

/transl\_table=11

/product="homoserine kinase"

/protein\_id="AA067737.1"

/db\_xref="GI:29136170"

/translation="MVKVYAPASSANMSVGFVLAAGAVTPVDGTLIGDVSVEAADHF  
RLHNLGRFADKLPPRENIVYQWERCQALGKTIPTVAMTLEKNMPTGSGLGSSACS  
VVAALVAMNEHCCKPLNDTRLALLAMELEGRISGSIHVDNVAFCPLGGNQLMIENGI  
ISQVPGFDEWLWLYAYPGIKVSTAEARAILPAQRQDCIAHGRHLAGFTIACYSRQ  
PLAALAMKMDVIAEPIYRARKLLPGFSQARQAVSEIGALASGISGSGPTLFCDCRKPETA  
QRVADWLKHYLQNGEGFVHICRLDTAGARVVG"

3734..5020

/gene="thrC"

/locus\_tag="t0004"

3734..5020

/gene="thrC"

/locus\_tag="t0004"

/note="corresponds to STY0004 from Accession AL513382:

Salmonella typhi CT18"

/codon\_start=1

/transl\_table=11

/product="threonine synthase"

/protein\_id="AA067738.1"

/db\_xref="GI:29136171"

/translation="NKLXNLKDHNEQVFAQAVTQGLGKQGLFPPHDLPEFSLTEID  
EMLNODFVSRSAKILSFIQDEIPQOILEERVAAFAFAPVAQVESDVGCLELPHGP  
TLAPKDEGRFMAQMLTHISGDKPVTILTATSGDTGAFAVAFAYLENVRVVILYPRG  
KISPLQEKLECTLGNIETVALDGFDAQALYKQAFDDEELKLTALGNSANISIR  
LLAQICYFEAVLQPGARQNLVTSVPSGNGDITAGLLAKSLKPLPVKRFIAATNIN  
DTQPTMRELKAGWATQATVSNAMDVSDQNNMPVEELFRKKIWRLTGELGYAAVDDT  
TIALPPEALAERADLPPLSHLHPADFAALRLKLMTRQ"

/gene="yaaA"

/locus\_tag="t0005"

complement(5114..5887)

/gene="yaaA"

/locus\_tag="t0005"

/note="corresponds to STY0005 from Accession AL513382:

Salmonella typhi CT18"

/codon\_start=1

/transl\_table=11

/product="conserved hypothetical protein"

/protein\_id="AA067739.1"

/db\_xref="GI:29136172"

/translation="MLILISPAKTLDYOSPLATTRYTOPELHDSQOLIQAROLASAP  
QISRLMGISDLADLNATRFHDWOPHTPDNARQAILAFKGDVYVTLGLOAFENDADFD  
FAQQHLRMLSGLYGLVRDLDMQPYRLEMGRLENPRCKDLYQFNGDITDKLNALE  
AQGRVVVNLAASEEYFKSVKPKKDAELIKPFDLDERNGKFKVSVFYAKKARGLMSRF  
ITENRLTKPELTAFDREGYFFDEETSTQDELVKRYEQ"

/gene="yaaJ"

/locus\_tag="t0006"

complement(5966..7396)

/gene="yaaJ"

/locus\_tag="t0006"

/note="corresponds to STY0006 from Accession AL513382:

Salmonella typhi CT18"

/codon\_start=1

/transl\_table=11

/product="putative amino-acid transport protein"

/protein\_id="AA067740.1"

/db\_xref="GI:29136173"

/translation="MPEFFSFINEILWGSVMYLLLLGAGCWFRTWRTGFIOFRIYIROFS  
RSLKGSLSQPQGLTSFQALCTSLAARIGSGNLGVALIAAGGPGVFMWVSAIIG

MATSFACSLAQYIKERDPTGQFRGGPANTWARGLMRWGVVFALELLVAYGLIFNS  
 VQNAVSRALHFAFNIPPLISGIALAFCALLIIIRIGIKGVARLMONLIPLIALIMVAG  
 SVFTCLWHIEQMPGVIASIVKSGWQEAAGAAGYLTQAITSGFORGMSFNEAGMG  
 STNAAATSYPPHPVAGQIVOMIGVSDTIIICITASAMILLAGHSHSSTEGIQ  
 LQHVMYLVGNGASGVALIVILFAFSSIVANYIYAENNLFFLRHLNHAIXWLLRLA  
 TLGMVIACTILISPLIQLADNMACHMAITNLTALILLSPPVITLAGDYLRQKRLGR  
 PQDPRFPDIEPOLAPDWTDAASRD

gene

7665..8618  
 /gene="tal8"  
 /locus\_tag="t0007"

CDS

7665..8618  
 /gene="tal8"  
 /locus\_tag="t0007"  
 /note="corresponds to STY0007 from Accession AL513382:  
 Salmonella typhi CT18"  
 /codon\_start=1  
 /transl\_table=11  
 /product="transaldolase B"  
 /protein\_id="AA067741.1"  
 /db\_xref="GI:29136174"

/translation="MTDKLSLRQFTTVVADTGDIAAMKLYQPDATNPISLIINAAQ  
 IPEYKLIDDAWAKAQSSDRAQOVVDATKLVNIGLEILKLVPGRISTEVDARLS  
 YDFASTAKAKRIKLYNDAGISNDRILIKLASTWQGIKRAEQLERKGINCLTILES  
 PAQARACAGVILISFPVGRILDWYKANDKKDPAEDPGVSVVTEIYEYKQHSY  
 ETVMGASFRNVEILELAGCDRLTIAPALLKELAESEGAIERKLSFSGEVKARPERI  
 TEAEFLWQHODPMVADKLADGIRKFAVDQEKLEKMGIDLL"

gene

8729..9319  
 /gene="mog"  
 /locus\_tag="t0008"

CDS

8729..9319  
 /gene="mog"  
 /locus\_tag="t0008"  
 /note="corresponds to STY0008 from Accession AL513382:  
 Salmonella typhi CT18"  
 /codon\_start=1  
 /transl\_table=11  
 /product="molybdopterin biosynthesis Mog protein"  
 /protein\_id="AA067742.1"  
 /db\_xref="GI:29136175"

/translation="MNTLRIGLVISDRASSGVVODKIPALEEWLASALTTPPELOT  
 RLIPDEQATIEQLCELVDMSCHLVLTGGTGPARDVTPDATLADRMVGFGEQ  
 MROISLRFVFTAILSRQGVIRKQALINLPQPKSIKRYETLEGVKADDDGVSVPFGIPA  
 SVPCIQLLDGPVETAPVAVPAPRPSARRENSS"

gene

/gene="yaah"  
 /locus\_tag="t0009"

CDS

complement(9376..9942)  
 /gene="yaah"  
 /locus\_tag="t0009"  
 /complement(9376..9942)  
 /gene="yaah"  
 /locus\_tag="t0009"  
 /note="corresponds to STY0009 from Accession AL513382:  
 Salmonella typhi CT18"  
 /codon\_start=1

## Alignment Scores:

Pred. No.: 2,27e-53 Length: 300169  
 Score: 990.00 Matches: 233  
 Percent Similarity: 64.18% Conservative: 59  
 Best Local Similarity: 51.21% Mismatches: 139  
 Query Match: 43.81% Indels: 24  
 DB: 1 Gaps: 10

US-09-701-229-2 (1-448) x AB016834 (1-300169)

Qy 2 SerLeuileAlaSerAspHisPheArg-----IleValValGlyLeuGlyLysSer 18  
 Db 148797 ACGTAATCATGCTGTAATACAGGACAAAACCTGCTATATFCGCTGGGCTTAAC 148856  
 Qy 19 GlyMetSerLeuValArgTyrLeuAlaArgGlyLeuProPheAlaValAlaAspThr 38  
 Db 148857 GGACTCTCTGCTGGACTTTTTCGCGCGCGCGTACGCGCGGGTGATGGATACT 148916  
 Qy .39 ArgGluAsnProGluLeuAlaThrLeuAlaGlnTyrProGln----ValGluVal 57  
 ||| ||||| ||| ||| ||||| |||||

Db 148917 CGCTGACGCCGCCGGCTCTGGATAAGCTG-----CGCAAGAGTTTGAGCGT 148964  
 Qy 58 ArgCysGlyGluLeuAspAlaGluPheLeuCysSerAlaArgGluLeuTyrValSerPro 77  
 Db 148965 CACGTTGGCGCCCTGAACGACGAGTGGCTCTTACGGCGGATTTAATGCTGCCAGCCCT 149024  
 Qy 78 GlyLeuSerLeuArgThrProAlaLeuValGlnAlaAlaLysGlyValArgLysSer 97  
 Db 149025 GGTATTGCTCTGGCGCATCCCTCTTAGCGCTCGCGCAGCGTGGAGTGAATATGTC 149084  
 Qy 98 GlyAspIleAspLeuPheAlaArgGluAlaLysAlaProIleValAlaIleThrGlySer 117  
 Db 149085 GCGCATATCGAATGTTTTCGCCGGAACGCGGCGGATGTGGCCATCACCGGCTCG 149144  
 Qy 118 AsnAlaLysSerThrValThrLeuValGlyGluMetAlaValAlaAlaAspLysArg 137  
 Db 149145 AACGGCAAAAGCAGCCGTGACCACTTAGTGGCGAGATGGCGAAGCGCGGCGTCAAT 149204  
 Qy 138 ValAlaValGlyGlyAsnLeuGlyThrProAlaLeuAspLeuLeuAlaAspIleGlu 157  
 Db 149205 GTCGGCGTTGGCGCAATATCGGCTTCCTCGCGCTGATGCTGCTGGATGCCGATCGCGAA 149264  
 Qy 158 LeuTyrValLeuGluLeuSerSerPheGlnLeuGluThrCysAspArgLeuAsnAlaGlu 177  
 Db 149265 TTGTACGTGCTGGAATATCCAGTTTCCAAATTGGAACCTACCTCAAGTTTTCGAACGCGCG 149324  
 Qy 178 ValAlaThrValLeuAsnValSerGluAspHisMetAspArgTyrAsp---GlyMetAla 196  
 Db 149325 CGGCAACGGTGTCAACGTCACATGATGATGACCGGTATCCGTTTGGTTGGAA 149384  
 Qy 197 AspTyrHisLeuAlaLysHisArgIlePheArgGlyAlaArgGlnValValAlaAsnArg 216  
 Db 149385 CAGTACCGCGCGGCGAAACTCGCGCTCTACGAGAAGCGAAGTGTGCTAGTGAATGCC 149444  
 Qy 217 AlaAspAlaLeuThrArgProLeuIleAlaAspThrValProCysTyrSerPheGlyLeu 236  
 Db 149445 GATCAGCGCTTGACGATCGCGGTACCGGGCCGATGAGCGCTCGTCAGCTTTGGCGTC 149504  
 Qy 237 AsnLysProAspPheLysAlaPheGlyLeuIleGluAspGlyGlnLysTrpLeuAla 256  
 Db 149505 AATATGGTGATTAATAC-----CTTAATCGTCAGAGGCGGGAACCTGCTACGA 149555  
 Qy 257 PheGlnPheAspLysLeuLeuProValGlyLeuLysIleArgGlyAlaHisAsnTyr 276  
 Db 149556 GTCAAGAGTGAGAGGTGCTGAACGTGAAGAGAGATGAAGCTTCCTCGGTACACATACTAT 149615  
 Qy 277 SerAsnAlaLeuAlaAlaLeuAlaLeuGlyHisAlaValGlyLeuProPheAspAlaMet 296  
 Db 149616 ACCAATCGTTAGCGCGCTGGCGGTGCGGATCCGTAGTCTGCCCGCGCCGACGAGT 149675  
 Qy 297 LeuGlyAlaLeuLysAlaPheSerGlyLeuAlaHisArgCysGlnTyrValArgGluArg 316  
 Db 149676 TTGAAGCGTTGACGACATTTTACCGGTCTGGCGATCGTTCAGTGGCGCTTGAGGAT 149735  
 Qy 317 GlnGlyValSerTyrAspSerLysAlaThrAsnValValGlyAlaLeuAlaAla 336  
 Db 149736 AACGGCGTCTGTTGGATCAACGACTCGAAAGCGACCAATGTCCGAGTACCGAAGCGCG 149795  
 Qy 337 IleGluGlyLeuGlyAlaAspIleAspGlyLysLeuValLeuLeuAlaGlyLysAspGly 356  
 Db 149796 CTAACGCGTTTG-----CATGTGGAGCGGTACGCTGCTGCTGCTGGCGGCGGACG 149849  
 Qy 357 LysGlyAlaAspPheHisAspLeuArgGluProValAlaArgPheCysArgAla----- 374  
 Db 149850 AAGTCGCGACACTTT-----TCTCCGCTGGCGGCTATCTGACCGCGGATGCT 149897  
 Qy 375 -----ValValLeuLeuGlyArgAspAlaGlyLeuIleAlaGlnAlaLeuGlyAsnAla 392  
 Db 149898 ATCCGCTGTATTGCTTTGGCGGATGGCGCGAGCTTGCG---GCACCTCGCTCGGAA 149954  
 Qy 393 ValProLeuValArgValAlaThrLeuAspGluAlaValArgGlnAlaAlaGluLeuAla 412  
 Db 149955 ATC---GCCCAACAGACTGAGACGATGGAAGCGGATGCGTTTGTCTGCTGCCGCCGCG 150011

QY	413	ArgGluclyAspAlaValLeuLeuSerProAlaCysAlaSerLeuLeuAspMetPheLysAsn	432
Db	150012	CACCCGGGAGATGCTGTGTGGCCGCCCTCGCCAGCCCGATCACTTTAAAAAT	150071
QY	433	PheGluclyArgGlyArgLeuPheAlaLysAlaValGluGluLeu	447
Db	150072	TTTGACCAACGGGCGATGCTCTTACCCTGCTGGCAGAGATTA	150116
RESULT	8		
LOCUS	AE008699	22348 bp	DNA linear BCT 23-APR-2003
DEFINITION	Salmonella typhimurium LT2, section 7 of 220 of the complete genome.		
ACCESSION	AE008699	AE006468	
VERSION	AE008699.1	GI:16418608	
KEYWORDS			
SOURCE			
ORGANISM	Salmonella typhimurium LT2		
REFERENCE	Salmonella typhimurium LT2		
AUTHORS	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Salmonella.		
	1 (bases 1 to 22348)		
	McClelland, M., Sanderson, K.E., Spieth, J., Clifton, S.W., Latreille, P., Courtney, L., Porwollik, S., Ali, J., Dante, M., Du, F., Hou, S., Layman, D., Leonard, S., Nguyen, C., Scott, K., Holmes, A., Grewal, N., Mulvaney, E., Ryan, E., Sun, H., Florea, L., Miller, W., Stoneking, T., Nhan, M., Waterston, R. and Wilson, R.K.		
TITLE	Complete genome sequence of Salmonella enterica serovar Typhimurium LT2		
JOURNAL	Nature	413 (6858), 852-856	(2001)
MEDLINE	21534948		
PUBMED	11677609		
REFERENCE	2 (bases 1 to 22348)		
AUTHORS	The Salmonella typhimurium Genome Sequencing Project		
CONSTRM	Direct Submission		
TITLE	Submitted (29-MAR-2001) Genome Sequencing Center, Department of Genetics, Washington University School of Medicine, 4444 Forest Park Boulevard, St. Louis, MO 63108, USA		
JOURNAL	Park Boulevard, St. Louis, MO 63108, USA		
COMMENT	COMMENT Supported by NIH grant 5U 01 AI43283		
	Coding sequences below are predicted from manually evaluated computer analysis, using similarity information and the programs: GLIMMER; <a href="http://www.tigr.org/softlab/glimmer/glimmer.html">http://www.tigr.org/softlab/glimmer/glimmer.html</a> and Genemark; <a href="http://opal.biology.gatech.edu/genemark/">http://opal.biology.gatech.edu/genemark/</a>		
	EC numbers were kindly provided by Junko Yabuzaki and the Kyoto Encyclopedia of Genes and Genomes; <a href="http://www.genome.ad.jp/kegg/">http://www.genome.ad.jp/kegg/</a> , and Pedro Romero and Peter Karp at EcoCyc; <a href="http://ecocyc.pangeasystems.com/ecocyc/">http://ecocyc.pangeasystems.com/ecocyc/</a>		
	The analyses of ribosome binding sites and promoter binding sites were kindly provided by Heladia Salgado, Julio Collado-Vides and <a href="http://kinich.cifn.unam.mx:8850/db/regulondb_intro.frameset">http://kinich.cifn.unam.mx:8850/db/regulondb_intro.frameset</a>		
	This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistries or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one ml3 subclone.		
FEATURES	Location/Qualifiers		
source	1..22348		
	/organism="Salmonella typhimurium LT2"		
	/mol_type="genomic DNA"		
	/strain="LT2; SGSC 1412; ATCC 700720"		
	/db_xref="ATCC:700720"		
	/db_xref="taxon:99287"		
	/note="LT2"		
	/complement(298..915)		
gene	/gene="leuB"		
	/note="synonym: STM0110"		

```

/transl_table=11
/product="2-isopropylmalate synthase"
/protein_id="AAL19077.1"
/db_xref="GI:16418612"
/translation="MSQOVIIFDTTLRDGEQALQASLSAKEKIQIALALERMGVDMVE
VGPVSPGDFESVQIARTIKNSRVCALARCVEKDIDVAQAQKLVADAFRIHTFIAT
SPMHIAIKLSTLDEIVIERAVYMKRARNYTDVEFSCEDAGRTPPVDGLARVVEAAIN
AGARTINPTDVTGYTPFFAGIISGLYERVPNIDKAIISVHTDDGLIAGVNSLAAY
HAGAROVGAMNIGIRGACALEEVIMAKVRKIDMNVHTNINHHIEIWRTSQTSOI
CNPPIANAIYVSGSAFAHSSGTHODGVKLNREYEMTPESIGLNOIQNLTSRSGR
AAYKRWEMGKDTYNDMDHLDAELKADKKGVDFDLEALAIINQKQEEPEHFR
LDYFSVQSSSDIATASVLACGEEIKAEAAAGNPVDIAIYQAINRTIGDYVELVKYD
LNAGKGKDALGQVDIVNHHGRFHVGLATDIVESSAKAMVHVLNNIWRAEVEKE
LQRAQKNKNNKETV"
RBS
complement(3417..3422)
/genes="leuB"
/notes="putative RBS for leuB; RegulonDB: STMS1H000492"
complement(4994..4999)
/genes="leuA"
/notes="putative RBS for leuA; RegulonDB: STMS1H000493"
complement(5067..5298)
/genes="leuL"
/notes="synonym: STM0114"
complement(5067..5153)
/genes="leuL"
/notes="leu operon leader peptide. (SW:LPL_SALTY)"
/codon_start=1
/transl_table=11
/product="leu operon leader peptide"
/protein_id="AAL19078.1"
/db_xref="GI:16418613"
/translation="MSHIVRFTGLLLNLFIVRGRVPGGIQH"
complement(5162..5167)
/genes="leuL"
/notes="putative RBS for leuL; RegulonDB: STMS1H000494"
complement(5271..5279)
/genes="leuL"
/notes="putative -10_signal for leuL; RegulonDB: STM1TH00472"
complement(5290..5298)
/genes="leuL"
/notes="putative -35_signal for leuL; RegulonDB: STM1TH00472"
5798..5754
/genes="leuO"
/notes="synonym: STM0115"
5798..5803
/genes="leuO"
/notes="putative RBS for leuO; RegulonDB: STMS1H000495"
5810..5754
/genes="leuO"
/notes="probable activator protein in leuabcd operon. (SW:LEUO_SALTY)"
/codon_start=1
/transl_table=11
/product="putative Lysr family transcriptional regulator"
/protein_id="AAL19079.1"
/db_xref="GI:16418614"
/translation="MPEVTEKPHLLDMGKPOLRMVDNLITVFDVMOEONITRAAH
TLGMSPAVSNVAKLVKMFNDLFRYRGIOPTARAFQIFGFSVKAOLQVONELPG
SGPEPTSSSRVNCVCSPLNDILTQIYNRVEKIAPNIHVPEKASLNQNTHEOLRYQ
ETEFVTSIEEPRPEFTSPLEFKDEMVLVSRKRPISGFLPGLDGVYNQEHAVYSLDR
YASFQSPWDTDPKQSVAYQGMALISVLNVQSQTHLVAIPRWLAEEFAESLDLIQL
PLPLKLSRTECYLSWHEAAGRDKGHWMDLLVSVCKR"
6941..6953
protein_bind
1.55e-54 Length: 22348
Pred. No.: 987.00 Matches: 232
Score: 63.96% Conservative: 59
Percent Similarity: 50.99% Mismatches: 140
Best Local Similarity: 43.67% Indels: 24
Query Match: 1 Gaps: 10
DB:

```

```

US-09-701-229-2 (1-448) x AE008699 (1-22348)
QY 2 SerLeuIleAlaSerAspHisPheArg-----IleValValGlyLeuGlyLysSer 18
Db 18632 ACCTTAATCATGCTGATTACCAGGACAAAACGCTGCTATTATGCTGGGCTTAAC 18691
QY 19 GlyMetSerLeuValArgTyrLeuAlaArgGlyLeuProPheAlaValAlaThr 38
Db 18692 GGACTCTCTGCGTGGACCTTTTCTCGCCGCGGCGTGACGCGCGGTGATGGACT 18751
QY 39 ArgGluAsnProGluLeuAlaThrLeuArgAlaGlnTyrProGln----ValGluVal 57
Db 18752 CGCGTCACGCGCGCGGTCTGGATAAGCTG-----CGCAAGAGGTGAGCGT 18799
QY 58 ArgCysGlyLeuLeuAspAlaGluPheLeuCysSerAlaArgGluLeuTyrValSerPro 77
Db 18800 CAGTTGGCGGCGCTGAACGACGAGTGGCTTTAGCGCGGATTTAATCGTCGCCAGCCCT 18859
QY 78 GlyLeuSerLeuArgThrProAlaLeuValGlnAlaAlaLysGlyValArgIleSer 97
Db 18860 GGTATTGCTCTGGCGCATCCCTCTCTTAGCGCTGCCCGCAGCGCTGGAGTGGTGC 18919
QY 98 GlyAspIleAspLeuPheAlaArgGluAlaLysAlaProIleValAlaIleThrGlySer 117
Db 18920 GCGGATATCGAAGTGTTCGCGAAGCGCGGATTTGCGCATCACCGGCTCG 18979
QY 118 AsnAlaLysSerThrValThrLeuValGlyGluMetAlaValAlaAlaAspLysArg 137
Db 18980 AACGCAAAAGCACCGCTGACCCCTTAGTGGCGGAGATGCGAAAGCGGCGGCGTCAAT 19039
QY 138 ValAlaValGlyGlyAsnLeuGlyThrProAlaLeuAspLeuLeuAlaAspIleGlu 157
Db 19040 GTCGCGGTGGCGCAATATCGCTTCGCGCGCTGATGCTGGTGGATCGCGGAA 19099
QY 158 LeuTyrValLeuGluLeuSerSerPheGlnLeuGluThrCysAspArgLeuAsnAlaGlu 177
Db 19100 TTGTAGCTGCTGGAAATATCCAGTTCCAACTGGAACCTACCTCAAGTTTTCGAAGCGCG 19159
QY 178 ValAlaThrValLeuAsnValSerGluAspHisMetAspArgTyrAsp---GlyMetAla 196
Db 19160 GCGGCAACGGTGTCAACGTCACCTGAAGATCATATGACCGGTATCCGTTGGTTGCCAA 19219
QY 197 AspTyrHisLeuAlaLysHisArgIlePheArgGlyAlaArgGlnValAlaAsnArg 216
Db 19220 CAGTACGCGGCGCGAAACCTGCGCTCTACGAGAGCGGAAAGTTCGCTAGTAGTGC 19279
QY 217 AlaAspAlaLeuThrArgProLeuIleAlaAspThrValProCysTyrSerPheGlyLeu 236
Db 19280 GATGACGCGTGTGACGATGCGGTACGCGGCGCGGATGAACGCTCGCTTAGCTTGGCGTC 19339
QY 237 AsnLysProAspPheLysAlaPheGlyLeuLeuGluAlaAspGlyGlnLysTrpLeuAla 256
Db 19340 AATATGGCGCATATCAC-----CTTAATCGTCAGCGGCGGAAACCTGCTCGCA 19390
QY 257 PheGlnPheAspLysLeuLeuProValGlyGluLeuLysIleArgGlyAlaHisAsnTyr 276
Db 19391 GTCAAAGCGCGAAGAGTGTGAACGTGAAGAGAGATGAAGCTTTCGCGTCAGCATACAT 19450
QY 277 SerAsnAlaLeuAlaAlaLeuAlaLeuGlyHisAlaValGlyLeuProPheAspAlaMet 296
Db 19451 ACCAATGCGTGTAGCGGCGTGGCGGCTGCGGATGCCGTAGCTCTCGCGCGCGCAGAGT 19510
QY 297 LeuGlyAlaLeuLysAlaPheSerGlyLeuAlaHisArgCysGlnTrpValArgGluArg 316
Db 19511 TTGAAGCGGTGACGACATTTACCGGTCTGCGCATCGCTTCAGCTGGCGGTGGAGCAT 19570
QY 317 GlnGlyValSerTyrTyrAspSerLysAlaThrAsnValGlyAlaAlaLeuAlaAla 336
Db 19571 AACGCGGTAGCTTGGATCAACGACTCGAAAGCGGACCAACCTCGCGCAGTACCGAGCGCG 19630
QY 337 IleGluGlyLeuGlyAlaAspIleAspGlyLysLeuValLeuLeuAlaGlyLysGlyAspGly 356

```

Dd 19631 CTAACGGTTG-----CATGTGGAGCTAGCCTACATCTGCTGCTGGCGGGCACGGT 19689

Qy 357 LysGlyAlaAspPheHisAspLeuArgGluProValAlaAraPhcCysArgAla----- 374  
||| ||||||| |||::: |||:::  
Db 19685 AACTCGGCAGACTTT-----TCCTCGCTGACACGGTATCTGACC GGCGATCGT 19732

Qy 375 -----valValLeuLeuGlyArgAspAlaGlyLeuIleAlaGlnAlaLeuLysAsnAla 392  
::: ||||||| |||::: |||:::  
Db 19733 ATCCGGCTGTATTGCTTTGGCGCGATGCCGCGACACTTGCC---GCACCTCGCTCCGAA 19789

Qy 393 valProLeuValArgValAlaThrIleuAspGluAlaValArgGlnAlaAglAlaLeuAla 412  
::: |||::: |||::: |||::: |||:::  
Db 19790 ATC---GCCCAACAGACTGAGAGCATGGAAGAGCGATCGCTTCTGCTCGCCGCGCGT 19846

Qy 413 ArgGluGlyAspAlaValLeuLeuSerProAlaCysAlaSerLeuAspMetPheLysAsn 432  
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 19847 CAGCCGGGTGATGTGCTGCTGTCGCCGCTTCGGCTAGCTCGATCAGTTAAAAAT 19906

Qy 433 PheGluGluGlyArgLeuPheAlaLysAlaValGluGluLeu 447  
|||||:::||||| ::||| :::: |||::: |||:::  
Db 19907 TTTGAGCAACGGGGCGATGCTTTACCCGCTGTCGCGAAGGAGTTA 19951

.RESULT 9

ECMUROY                2608 bp     DNA     linear     BCT 12-SEP-1993

LOCUS

DEFINITION      Escherichia coli murD gene and ORF-Y (EC 6.3.2.9).

ACCESSION      X51584.X52117

VERSION        X51584.1 GI:42058

KEYWORDS       membrane protein; murD gene;

                UDP-N-acetylmuramoyl-L-alanyl-D-glutamate synthetase; unidentified reading frame.

SOURCE

ORGANISM       Escherichia coli

Bacteria: Proteobacteria; Gammaproteobacteria; Enterobacteriales;

Enterobacteriaceae; Escherichia.

REFERENCE

AUTHORS       Ikeda,M., Wachi,M., Ishino,F. and Matsubashi,M.

TITLE          Nucleotide sequence involving murD and an open reading frame ORF-Y spacing murF and ftsw in Escherichia coli

JOURNAL       Nucleic Acids Res. 18 (4), 1058 (1990)

MEDLINE       90192099

PUBMED        2179861

REFERENCE

AUTHORS       Ikeda,M.

TITLE          Direct Submission

JOURNAL       Submitted (02-JAN-1989) Ikeda M., Institute of Applied Microbiology, The University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113, Japan

FEATURES

1..2608

/organism="Escherichia coli"

/mol\_type="genomic DNA"

/db\_xref="taxon:562"

/map="2min"

98..102

/note="pot. ribosome binding site"

111..1193

/note="ORF-Y (AA 1-360)"

/codon\_start=1

/transl\_table=11

/protein\_id="CAA35932.1"

/db\_xref="GI:42059"

/db\_xref="SWISS-PROT:P15876"

/translation="MLWVAHLVKKYSGNFVSYLTFRAIVSLLTALFISLWMGPRLTAHQKLSFGQVRNDPESFSKRGTPTMGIGILTAIVISLVLLNAYPSNPYYWCVL

RBS

CDS

VVLVGVIGVIVDDYVRKDDTKGLIARWKYEFWVSIALGVAFALYLAGKOTPATQIQ  
VWFFKQVMPQLGFLYITLLAYFIVIVGTGNAVNLTDGLDGLAIIMPTVFAVGGFALVWA  
TGNNFASYLHPIYLRHAGELVITCVTAIVGAGLGFLEWNTYPAQVFMGDVGSIALGGA  
LGIIVALLROEFLIVTMGGVFVETLSVILQVGSFKLGRQIRPRMAPIHHHYELAKGW  
EPRIVVRFIISLMLVLIGLTKVR"

promoter 1018..1023  
/note="pot. -35 region"  
promoter 1043..1048  
/note="pot. -10 region"  
RBS 1181..1185  
/note="pot. ribosome binding site"  
CDS 1196..12512  
/note="murd gene product (AA 1-438)"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAA35933.1"  
/db\_xref="GI:42060"  
/db\_xref="SWISS-PROT:P14900"  
/translation="MADYQCKNVIIVIGLITGLSCVDFFLARGVTPRVMDTRMTPPGL  
DKLPEAVERTHTSGINDEWLMADILVASQIALAHSPLSAAADAGITVDGIELVFLCS  
AQAPIVAITCSNGKSTVTLVIGEMAKAAGVNVGVGNIGLPAMLMLDDECELYELFELS  
SFOLETSSIQAAVATLANTVEDHMDRYPEGLQOYRAAKLRIYENAKVCVNVADDAIT  
MPRTGADERCVSGVNGMDYHLNHQOGETWLRVKGKLVNEMKLSGHHNTNALAA  
LALADAGLPRASLKAALTFTTGLPHRFVEVLEHNGVRINDSKSNATVNSVEATEANGL  
HVPDTHLLGGGKGADEFLARYLNGDNVRLYCFGRDQAQLAALRPVEVAGTETME  
QAMRLAPRVQPDGMWLLSPACASLDQFKNFQEGRNEFARLAKELG"

BASE COUNT 530 a 631 c 777 g 670 t  
ORIGIN

Alignment Scores:  
Pred. No.: 1.35e-55 Length: 2608  
Score: 986.00 Matches: 230  
Percent Similarity: 63.96% Conservative: 61  
Best Local Similarity: 50.55% Mismatches: 140  
Query Match: 43.63% Indels: 24  
DB: 1 Gaps: 10

US-09-701-229-2 (1-448) x ECMUROY (1-2608)

Qy 2 SerLeuileAlaSeraspHisPheArg-----lleValValGlyLeuGlyLySer 18  
Db 1187 ACGTAAATCATGCTGATTAATCATAGGCTAAAAATGTCTCATTTATCGCCCTGGCCCTCAC 1246  
Qy 19 GlyMetSerLeuValArgTyrLeuAlaArgArgGlyLeuPropheAlaValAlaAspThr 38  
Db 1247 GGCGTTTCCTGCGTGGACCTTTTCCCTCGCTCGCGGTGTGACGCCGCCGTATGGATACG 1306  
Qy 39 ArgGluasnProProGluLeuAlaThrLeuArgAlaGlnTyrProGln---ValGluVal 57  
Db 1307 CGTATGACACCGCTGGCCCTGGATAAATTA-----CCCGAAGCGGTAGAAGCG 1354  
Qy 58 ArgCysGlyGluLeuAspAlaGluPheLeuCysSerAlaArgGluLeuTyrValSerPro 77  
Db 1355 CACACGGGCAGCTCGAATGATGAATCGCTGATGGCGGCAGATCTGTGATTCGCCAGTCCC 1414  
Qy 78 GlyLeuSerLeuArgThrProAlaLeuValGlnAlaAlaAlaLysGlyValArgIleSer 97  
Db 1415 GGTATTGCATGGCGCATCCATCCTTAAGCGCTGCCCTGATGCCGATGCCGAATCGAATCGTT 1474  
Qy 98 GlyAspIleAspLeuPheAlaArgGluAlaLysAlaProIleValAlaIleThrGlySer 117  
Db 1475 GCGGATATCAGCTGTTCTGTCGCGAAGCACAGCACCGCATGTGTGCGGATTAACGGTTCT 1534  
Qy 118 AsnAlaLysSerThrValThrLeuValGlyGluMetAlaValAlaAlaAspLysArg 137  
Db 1535 AACCGGAAAAGCAGCTCACAGCTAGTGGGTGAATGCCGAAAGCGCGGGGTAAAC 1594  
Qy 138 ValAlaValGlyGlyAsnLeuGlyThrProAlaLeuAspLeuAlaAspAspIleGlu 157  
Db 1595 GTTGGTGTGGGTGGCAATATATGGCCCTGCTTACTGTGATGATGAGTGTGTA 1654  
Qy 158 LeuTyrValLeuGluLeuSerSerPheGlnLeuGluThrCysAspArgLeuAsnAlaGlu 177





```

/note="factor Sigma70; promoter leu; documented +1
at83735"
protein_bind
complement(201..221)
/note="No predicted promoter"
/bound_moiety="LexA predicted site"
protein_bind
complement(201..221)
/note="central position to predicted promoter: -0.5"
/bound_moiety="LexA predicted site"
protein_bind
complement(354..368)
/note="No predicted promoter"
/bound_moiety="Lrp documented site"
protein_bind
complement(354..368)
/note="central position to leuABCD promoter: -156.5"
/bound_moiety="Lrp documented site"
protein_bind
460..474
/note="central position to predicted promoter: -28.5"
/bound_moiety="Lrp predicted site"
promoter
462..490
/note="factor Sigma70; predicted +1 start at 84042"
complement(519..537)
protein_bind
/note="central position to predicted promoter: -319.5"
/bound_moiety="TyrR predicted site"
protein_bind
554..565
/note="central position to predicted promoter:64"
/bound_moiety="AraC predicted site"
gene
646..1767
/note="leuO"
/note="synonym: b0076"
CDS
646..1767
/note="leuO"
/function="putative regulator; Amino acid biosynthesis:
Leucine"
/note="o373; 100 pct identical to LEUO_ECOLI SW:
p10151(290 aa) but contains 59 additional N-ter aa and 24
C-term residues; 100 pct identical to PIR: S40589 but
contains 24 additional C-term residues"
/codon_start=1
/transl_table=11
/product="probable transcriptional activator for leuABCD
operon"
/protein_id="AAC73187.1"
/db_xref="GI:1786264"
/transl_table=11
/note="central position to predicted promoter: -39.5"
/bound_moiety="Lrp predicted site"
promoter
1813..1839
/note="factor Sigma70; promoter ilvIhp4; documented+1 at
85394"
protein_bind
1833..1847
/note="central position to predicted promoter: -8.5"
/bound_moiety="Lrp predicted site"
promoter
1838..1866
/note="factor Sigma70; promoter ilvIhp3; documented+1 at
85420"
protein_bind
1915..1929
/note="central position to predicted promoter:73.5"
/bound_moiety="Lrp predicted site"
protein_bind
1949..1963
/note="central position to predicted promoter:107.5"
/bound_moiety="Lrp predicted site"
promoter
1955..1984
/note="factor Sigma70; promoter ilvIhp2; documented+1 at
85534"
protein_bind
1978..1992
/note="central position to predicted promoter:136.5"
/bound_moiety="Lrp predicted site"

```

```

1995..3809
/note="ilvI"
/note="synonym: b0077"
1995..3809
/note="ilvI"
/EC_number="4.1.3.18"
/function="enzyme; Amino acid biosynthesis: Isoleucine,
Valine"
fromILVI_ECOLI SW: P00893 (566 aa)
/codon_start=1
/transl_table=11
/product="acetolactate synthase III, valine sensitive,
large subunit"
/protein_id="AAC73188.1"
/db_xref="GI:1786265"
/transl_table=11
/note="central position to predicted promoter:156.5"
/bound_moiety="Lrp predicted site"
2017..2045
/note="ilvI"
/note="factor Sigma70; promoter ilvIhp1; documented+1 at
85597"
3812..4303
/note="ilvH"
/note="synonym: b0078"
3812..4303
/note="ilvH"
/EC_number="4.1.3.18"
protein_bind
/note="central position to predicted promoter:156.5"
/bound_moiety="Lrp predicted site"
2017..2045
/note="ilvI"
/note="factor Sigma70; promoter ilvIhp1; documented+1 at
85597"
3812..4303
/note="ilvH"
/note="synonym: b0078"
3812..4303
/note="ilvH"
/EC_number="4.1.3.18"
Alignment Scores:
Pred. No.: 1.74e-54 Length: 21757
Score: 986.00 Matches: 230
Percent Similarity: 63.96% Conservative: 61
Best Local Similarity: 50.55% Mismatches: 140
Query Match: 43.63% Indels: 24
DB: 1 Gaps: 10
US-09-701-229-2 (1-448) x AE000118 (1-21757)
Qy 2 SerLeuileAlaSerAspHisPhearg-----ileValValGlyLeuGlyLysSer 18
Db 13533 ACGTTAATCATGGCTGATGATATACAGGTTAAATATGCGTCATATCGGCTCGGCTCAC 13592
Qy 19 GlyMetSerLeuValArgTyrLeuAlaArgGlyLeuProPheAlaValValAspThr 38
Db 13593 GGGCTTCTCGCTGGACTTTTCTCGCTCGGCTGTGACGCGCGCTTATGGATACG 13652
Qy 39 ArgGluAsnProProGluLeuAlaThrLeuArgAlaGlnTyrProGln---ValGluVal 57
Db 13653 CGTATGACACCGCTCGCTGATAAATA-----CCGAAAGCGGTAGAACGC 13700
Qy 58 ArgCysGlyGluLeuAspAlaGluPheLeuCysSerAlaArgGluLeuTyrValSerPro 77
Db 13701 CACAGGGCAGCTCTGAATGATGATGCTGATGGCGGCGAGATCTGATGTCGCCAGTCCC 13760
Qy 78 GlyLeuSerLeuArgThrProAlaLeuValGlnAlaAlaLysGlyValArgIleSer 97
Db 13761 GGTATTGACGTGGCGCATCCATCCTTAAGCGCTCGCGCTGATCGCGGAATCGAATCGTT 13820
Qy 98 GlyAspIleAspLeuPheAlaArgGluAlaLysAlaProIleValAlaIleThrGlySer 117

```





```
Db 14196 CAGTATCGCGAGCAAACTGCGCATTTACGAAACGCGAAAGTTTGCCTGTTAATGCT 14255
QY 217 AlaAspAlaLeuThrArgProLeuIleAlaAspThrValProCysTrpSerPheGlyLeu 236
Db 14256 GATGATGCTTTACAAATGCGGATTCGGTGGCGGATGAACGCTGCGTCACTTGGCGTC 14315
QY 237 AsnLysProAspPheLysAlaPheGlyLeuIleGluAspGlyClnLysTrpLeuAla 256
Db 14316 AACATGGGTGACTATCAC-----CTGAATCATCAGCGCGCAACCTGGCTGCGG 14366
QY 257 PheGlnPheAspLysLeuLeuProValGlyGluLeuLysIleArgGlyAlaHisAsnTrp 276
Db 14367 GTTAAAGCGGAGAAAGTGTGAATGTAAGAGATGAACACTTTCGGGCGAGCATAACTAC 14426
QY 277 SerAsnAlaLeuAlaLeuGlyLysAlaValGlyLeuProPheAspAlaMet 296
Db 14427 ACCAATGCGCTGGCGGCTGGCGTGGCGAGATGTCGAGGCTTACCGGTCGCCAGCAGC 14486
QY 297 LeuGlyAlaLeuLysAlaPheSerGlyLeuAlaHisArgCysGlnTrpValArgGluArg 316
Db 14487 CTGAAGCGCTTAACACCATCTACTGCTGCGCATCGCTTTGAAGTTGCTGGGCGCAT 14546
QY 317 GlnGlyValSerTyrThrAspSerLysAlaThrAsnValGlyLeuAlaLeuAla 336
Db 14547 AACGGCGTACGTTGGATTACGATTGCAAGAGCGACCACTGCGGAGTACGGAACGGCGG 14606
QY 337 IleGluGlyLeuGlyAlaAspIleAspGlyLysLeuValLeuAlaGlyGlyAspGly 356
Db 14607 CTGAATGGCGCTG-----CACGTAGACGCGACACTGCTTTGCTGGTGGCGATGCT 14660
QY 357 LysGlyAlaAspPheHisAspLeuArgGluProValAlaAlaArgPheCysArgAla----- 374
Db 14661 AAATCGCGGAGCTTT-----AGCCCACTGGCGGCTTACTGAATGGCGATAAC 14708
QY 375 -----ValValLeuGlyArgAspAlaGlyLeuIleAlaGlnAlaLeuGlyAsnAla 392
Db 14709 GTACGCTCTATTTGCTGCGTGGCGGCGGCGAGCTGGCG--CGCGTACGCCCGGAA 14765
QY 393 ValProLeuValArgValAlaThrLeuAspGlyAlaValArgGlnAlaGluLeuAla 412
Db 14766 GTG---GCAGAACAAACCGAACTATGGAACAGCGGATGCGCTTGTGCTCCGCTGTT 14822
QY 413 ArgGluGlyAspAlaValLeuLeuSerProAlaCysAlaSerLeuAspMetPheLysAsn 432
Db 14823 CAGCGCGGCGGATGTTGCTCTCCCGAGCTGTGCCAGCTTGCATCAGTTCAAGAAC 14882
QY 433 PheGluGluArgGlyArgLeuPheAlaLysAlaValGluGluLeu 447
Db 14883 TTGTAACACGAGGCAATGATGTTGCCGCTCTGGCGAAGGAGTTA 14927

RESULT 12
EC2MIN
ID BC2MIN standard; DNA; PRO; 28277 BP.
XX
AC X55034; M10429;
XX
SV X55034.1
XX
DT 21-FEB-1991 (Rel. 27, Created)
DT 05-JUL-1999 (Rel. 60, Last updated, Version 31)
DE E. coli 2 minute region
XX
KW dcl gene; envA gene; ftsA gene; ftsQ gene; ftsW gene; ftsZ gene; ilvH gene;
KW ilvI gene; leuA gene; leuO gene; mraY gene; murC gene; murD gene;
KW murE gene; murF gene; murG gene; mutR gene; orfA; orfB; orfC; orfX;
KW pbpB gene; secA gene; shl gene.
OS Escherichia coli
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
XX
```

[1]  
RN 1-28277  
RP Ayala J.A.;  
RT Submitted (08-JAN-1991) to the EMBL/GenBank/DBJ databases.  
RL Ayala J.A., Instituto de Biologia Molecular, Centro de Biologia Molecular,  
RL Universidad Autonoma, Canto-Blanco 28049, Madrid, Spain.  
XX

[2]  
RN Ayala J.A.;  
RA "Regulation of transcription at the 2-minute region of the genetic map of  
RT Escherichia coli";  
RL Unpublished.  
XX

[3]  
RN 1-306  
RP MEDLINE; 82078077.  
RX PUBMED; 6171647.  
RA Wessler S.R., Calvo J.M.;  
RT "Control of leu operon expression in Escherichia coli by a transcription  
RT attenuation mechanism";  
RL J. Mol. Biol. 149(4):579-597(1981).  
XX

[4]  
RN 268-1130  
RP MEDLINE; 86223773.  
RX PUBMED; 3519576.  
RA Haughn G.W., Wessler S.R., Gemmill R.M., Calvo J.M.;  
RT "High A + T content conserved in DNA sequences upstream of leuABCD in  
RT Escherichia coli and Salmonella typhimurium";  
RL J. Bacteriol. 166(3):1113-1117(1986).  
XX

[5]  
RN 843-1812  
RP MEDLINE; 88320486.  
RX PUBMED; 3413113.  
RA Henikoff S., Haughn G.W., Calvo J.M., Wallace J.C.;  
RT "A large family of bacterial activator proteins";  
RL Proc. Natl. Acad. Sci. U.S.A. 85(18):6602-6606(1988).  
XX

[6]  
RN 1799-2187  
RP MEDLINE; 85234358.  
RX PUBMED; 3891724.  
RA Haughn G.W., Squires C.H., Defelice M., Largo C.T., Calvo J.M.;  
RT "Unusual organization of the ilvIH promoter of Escherichia coli";  
RL J. Bacteriol. 163(1):186-198(1985).  
XX

[7]  
RN 2101-4431  
RP MEDLINE; 83272971.  
RX PUBMED; 6308579.  
RA Squires C.H., Defelice M., Devereux J., Calvo J.M.;  
RT "Molecular structure of ilvIH and its evolutionary relationship to ilvG in  
RT Escherichia coli";  
RL Nucleic Acids Res. 11(15):5299-5313(1983).  
XX

[8]  
RN 4274-6093  
RP MEDLINE; 90330585.  
RX PUBMED; 2198273.  
RA Leclerc G., Noel G., Drapeau G.;  
RT "Molecular cloning, nucleotide sequence and expression of shl, a new gene  
RT in the 2-minute region of the genetic map of Escherichia coli";  
RL J. Bacteriol. 172(8):4696-4700(1990).  
XX

[9]  
RN 6088-7587  
RP MEDLINE; 90251464.  
RX Gomez M.J., Fluoret B., Van Heijenoort J., Ayala J.A.;  
RT "Nucleotide sequence of the regulatory region of pbpB gene of Escherichia  
RT coli";  
RL Nucleic Acids Res. 18:2813-2813(1990).



[illegible]

[illegible]



```
Db 316 CCGATTGTGGCGATTACCGGTTCTAAACGGCAAAAGACGGTCACCGACTAGTGGTGAA 375
Qy 130 MetAlaValAlaAlaAspLysArgValAlaValGlyGlyAsnLeuGlyThrProAlaLeu 149
Db 376 ATGGCGAAGCGCGGGTTAACTGGTGGTGGTGGCAATATGGCCCTGCCCTGCGGTG 435
Qy 150 AspLeuLeuAlaAspAspIleGluLeuTyrValLeuGluLeuSerSerPheGlnLeuGlu 169
Db 436 ATGCTACTGGATGATGAGTGTGAACCTGTACGTGTGGAACCTGTGAGCTTCCAGCTGGAA 495
Qy 170 ThrCysAspArgLeuAsnAlaGluValAlaThrValLeuAsnValSerGluAspHisMet 189
Db 496 ACCACCTCCAGCTTACAGCGGTAGCAGCGACCATCTCTGAACGTGACTGAAGATCATATG 555
Qy 190 AspArgTyrAsp---GlyMetAlaAspTyrHisLeuAlaLysHisArgIlePheArgGly 208
Db 556 GATCGCTATCGTTGGTTTACACAGTATCGTCAGCAAAACCTGCGCATTTACAAAC 615
Qy 209 AlaArgGlnValValAlaAsnArgAlaAspAlaLeuThrArgProLeuIleAlaAspThr 228
Db 616 CGCAAAAGTTTCGCTGGTAAATGCTGATGCCCTTAACAATGCCGATTCGCGGTGCGGAT 675
Qy 229 ValProCysTrpSerPheGlyLeuAsnLysProAspPheLysAlaPheGlyLeuIleGlu 248
Db 676 GAACGCTGCGTCAGCTTGGCGCTCAACATCGCTGGGTGACTATCAC-----CTGAATCAT 726
Qy 249 GluAspGlyGlnLysTrpLeuAlaPheGlnPheAspLysLeuLeuProValGlyGluLeu 268
Db 727 CAGCAGGCGAAACCTGGCTCGGGTTAAAGCGCGAAGAGTGCATGTGAAGAGATG 786
Qy 269 LysIleArgGlyAlaHisAsnTyrSerAsnAlaLeuAlaLeuAlaLeuGlyHisAla 288
Db 787 AAATCTCCGGCAGCATACTACACCAATCGCTGGCGCGCTGGCGCTGGCAGATGCT 846
Qy 289 ValGlyLeuProPheAspAlaMetLeuGlyAlaLeuLysAlaPheSerGlyLeuAlaHis 308
Db 847 GCAGGTTACCGCTGCCAGCGCTGAACCGGTAAACCATTCATCTGCTGCCGAT 906
Qy 309 ArgCysGlnTrpValArgGluArgGlnGlyValSerTyrTyrAspSerLysAlaThr 328
Db 907 CGCTTTGAAGTTGTGGGACCATACCGCGCTACGTGGATTACGATTCAAGTTCGAAGCGACC 966
Qy 329 AsnValGlyAlaAlaLeuAlaAlaIleGluGlyLeuGlyAlaAspIleAspGlyLysLeu 348
Db 967 AACCTCGCGCATAGCGAAGCGCGCTGAATGGCCTG-----CACGTAGACGGCACACTG 1020
Qy 349 ValLeuLeuAlaGlyGlyAspGlyLysGlyAlaAspPheHisAspLeuArgGluProVal 368
Db 1021 CATTGTGTCTGGTGGCGATGGTAAATCGCGCGACTTT-----AGCCCACTG 1068
Qy 369 AlaArgPheCysArgAla-----ValValLeuLeuGlyArgAspAlaGlyLeu 384
Db 1069 CGCGGTTACCTGAATGCGGATAACGTACTGTATGTTTCGTCGTGACGGCGCGCAG 1128
Qy 385 IleAlaGlnAlaLeuGlyAsnAlaValProLeuValArgValAlaThrLeuAspGluAla 404
Db 1129 CTGGCG---GGCTACGCGCCGGAAGTG---GCAGCAAAACCGAAGTATGGAACAGCGC 1182
Qy 405 ValArgGlnAlaAlaGluLeuAlaArgGluGlyAspAlaValLeuLeuSerProAlaCys 424
Db 1183 ATGGCGTTGTGGTTCGCGGTTCAGCGCGGCGCATATGGTTCTGCTCTCCCGACCTGT 1242
Qy 425 AlaSerLeuAspMetPheLysAsnPheGluGluArgGlyArgLeuPheAlaLysAlaVal 444
Db 1243 GCCAGCTTGATCAGTTCAAGAACTTGAACACAGCGCATATGATTTGCCCTCTGGCG 1302
Qy 445 GluGluLeu 447
Db 1303 AAGGAGTTA 1311
```

RESULT 15  
AE005185

```
LOCUS AE005185 12791 bp DNA linear BCT 21-MAR-2001
DEFINITION Escherichia coli O157:H7 EDL933 genome, contig 1 of 3, section 9 of
155.
ACCESSION AE005185 AE005174
VERSION AE005185.1 GI:12512782
KEYWORDS
SOURCE Escherichia coli O157:H7 EDL933
ORGANISM Escherichia coli O157:H7 EDL933
REFERENCE 1 (bases 1 to 12791)
AUTHORS Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamouis,K.,
Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
Welch,R.A. and Blattner,F.R.
TITLE Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
JOURNAL Nature 409 (6819), 529-533 (2001)
MEDLINE 21074935
PUBMED 11206551
REFERENCE 2 (bases 1 to 12791)
AUTHORS Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamouis,K.,
Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
Welch,R.A. and Blattner,F.R.
TITLE Direct Submission
JOURNAL Submitted (22-OCT-2000) Laboratory of Genetics, University of
Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
FEATURES
source
1..12791
/organism="Escherichia coli O157:H7 EDL933"
/mol_type="genomic DNA"
/strain="EDL933"
/serotype="O157:H7"
/db_xref="taxon:155864"
/note="enterohaemorrhagic"
72..1076
/gene="fruF"
/note="synonym: Z0090"
72..1076
/gene="fruF"
/function="regulator; Energy metabolism, carbon:
Glycolysis"
/note="Residues 1 to 334 of 334 are 100.00 pct identical
to residues 1 to 334 of 334 from Escherichia coli K-12
Strain MG1655: B0080"
/codon_start=1
/transl_table=11
/product="transcriptional repressor of fru operon and
others"
/protein_id="AAG54384.1"
/db_xref="GI:12512783"
/translation="MKLDEIARLAGSVRTTASYVINGKAKQYRVSDTKTEVKVMVYRE
HNYHPNVAAGLRAGTRISIGLVPDLENTYRIANLYLRQARGQYQLIACSEDO
PDNEMRCLEHLQROVDAIIVSTLPPHPFQWANDPFPVIALDRALDREHFTSVV
GADODAEMLAEELKPEAFYLVYLGALPELSVFLREQGFATKWDKDPREHFTLVAN
SYEREAAQLAEKLETHPQALFTTTSFALLQGMVTLRDGKLPDSDLAIFGDN
ELLDFLQCPVLAAQRHDAERVLEIVLASLDEPRKPKPGLTRIKRLNRYRGLVRS
"
1678..2136
/gene="yabB"
/note="synonym: Z0091"
1678..2136
/gene="yabB"
/function="orf; Unknown function"
/note="Residues 1 to 152 of 152 are 99.34 pct identical to
residues 1 to 152 of 152 from Escherichia coli K-12 Strain
MG1655: B0081"
/codon_start=1
/transl_table=11
```







```
Qy 39 ArgGluAsnProProGluLeuAlaThrLeuArgAlaGlnTyrProGln---ValGluVal 57
Db 9242 CGTATGACACCGCTGGCTGGATAAATTA-----CCGGAAGCGTAGAAGC 9289
Qy 58 ArgCysGlyGluLeuAspAlaGluPheLeuCysSerAlaArgGluLeuTyrValSerPro 77
Db 9290 CACACGGCGAGTCTGAATGATGATGGCTGATGGCGCAGATCTGATGTCGCCAGTCCC 9349
Qy 78 GlyLeuSerLeuArgThrProAlaLeuValGlnAlaAlaLysGlyValArgIleSer 97
Db 9350 GGTATTGCACCTGGCGCATCCATCTTAAGCGCTGCCGCTGATGCCGGAATCGAATCGTT 9409
Qy 98 GlyAspIleAspLeuPheAlaArgGluAlaLysAlaProIleValAlaIleThrGlySer 117
Db 9410 GCGGATATCGAGCTGTCTCTGCGGAAGCACACGACCGGATTTGTGGGATTACCGGTTCT 9469
Qy 118 AsnAlaLysSerThrValThrLeuValGlyGluMetAlaValAlaAlaAspLysArg 137
Db 9470 AACGGCAAAAGCAGCGTCACCGCTAGTGGTGAATGGCGAAAGCGGGGGTTAAC 9529
Qy 138 ValAlaValGlyGlyAsnLeuGlyThrProAlaLeuAspLeuLeuAlaAspIleGlu 157
Db 9530 GTTGGTGTGGTGCAATATTGGCTGCCGCGTTGATGCTGCTGGATGATGAGTGGGAA 9589
Qy 158 LeuTyrValLeuGluLeuSerSerPheGlnLeuGluThrCysAspArgLeuAsnAlaGlu 177
Db 9590 CTGTACGTACTGGAACTGTCCAGCTTCCAGCTGGAAACCCCTCCAGCTTACAGCGGTA 9649
Qy 178 ValAlaThrValLeuAsnValSerGluAspHisMetAspArgTyrAsp---GlyMetAla 196
Db 9650 GCACGACCATCTGACGCTGACGAGATGATGATGATGATGATGATGATGATGATGATG 9709
Qy 197 AspTyrHisLeuAlaLysHisArgIlePheArgGlyAlaArgGlnValValAsnArg 216
Db 9710 CAATACCGTCGGCGCAAAATTCGTTATTACGAAATGCGAAAGTTTGTGCTCAATGCT 9769
Qy 217 AlaAspAlaLeuThrArgProLeuIleAlaAspThrValProCysTyrSerPheGlyLeu 236
Db 9770 GATGATGCTTTAACAATGCCGATTCGCGGTGCGGATGAACGCTGCGTCAGCTTTGGCGTC 9829
Qy 237 AsnLysProAspPheLysAlaPheGlyLeuIleGluAspGlyGlnLysTyrLeuAla 256
Db 9830 AACATGGGTGACTATCAC-----CTGATCATCAGCAGCGCGGAAACCTGGCTGGG 9880
Qy 257 PheGlnPheAspLysLeuLeuProValGlyLeuLysIleArgGlyAlaHisAsnTyr 276
Db 9881 GTGAAAGCGGAGAAAGTGTCAACGCTGAACAGAAATGAACCTTCCGGGACAGATACTAC 9940
Qy 277 SerAsnAlaLeuAlaAlaLeuAlaLeuGlyHisAlaValGlyLeuProPheAspAlaMet 296
Db 9941 ACCAATGCGTGGCGCGCTGGCGCTGCGAGATCCCGAGGTTTACCGCGCTGCCAGCAGC 10000
Qy 297 LeuGlyAlaLeuLysAlaPheSerGlyLeuAlaHisArgCysGlnTyrValArgGluArg 316
Db 10001 CTGAACGGTTAACCATTCATCTGCGCTGCGGCTGCTTGAAGTTGTGCTGGAGCAT 10060
Qy 317 GlnGlyValSerTyrTyrAspSerLysAlaThrAsnValGlyAlaAlaLeuAlaAla 336
Db 10061 AACGCGGTACGTTGATTAAGATTTCGAACGACCAACCGTCGCGAGTACGGAAGCGCG 10120
Qy 337 IleGluGlyLeuGlyAlaAspIleAspGlyLysLeuValLeuLeuAlaGlyGlyAspGly 356
Db 10121 CTGAATGGCTG-----CAGTAGAGCGCAGCTGCTGCTTTGCTGGCGCGGATGGT 10174
Qy 357 LysGlyAlaAspPheHisAspLeuArgGluProValAlaArgPheCysArgAla----- 374
Db 10175 AAATCGCGGACTTT-----AGCCACTGGCGGCTTACCTCAATGGCGATAAC 10222
Qy 375 -----ValValLeuLeuGlyArgAspAlaGlyLeuIleAlaGlnAlaLeuGlyAsnAla 392
Db 10223 GTACGTCGTGATTGTTGCGTCTGAC---GGCGTGAGCTGGCGCGCTACGCCCGGAA 10279
Qy 393 ValProLeuValArgValAlaThrLeuAspGluAlaValArgGlnAlaAlaGluLeuAla 412
```

```
Db 10280 GTG---GCAGAACAAACCGAAACTATGGAACAGCGGATGCGCTTGCTGCGCTGTT 10336
Qy 413 ArgGluGlyAspAlaValLeuLeuSerProAlaCysAlaSerLeuAspMetPheLysAsn 432
Db 10337 CAGCCGGCGGATATGGTTCTACTCTCCCGGCGCTGTGCCAGCCTTGATCAGTTCAGA 10396
Qy 433 PheGluGluArgGlyArgLeuPheAlaLysAlaValGluGluLeu 447
Db 10397 TTTCAACAACGAGCAATGAGTTTGCCCGCTGTCGCGAAGGAGTTA 10441
```

Search completed: August 14, 2003, 08:41:41  
Job time : 4825 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2003, 03:24:05 ; Search time 5305 Seconds  
(without alignments)  
11181.708 Million cell updates/sec

Title: US-09-701-229-1  
Perfect score: 1450  
Sequence: 1 cgtgctgctgcctgcga.....tgttgagcgcgcacggcgcac 1450

Scoring table: OLIGO.NUC  
Gapop 60.0 , Gapext 60.0

Searched: 2888711 seqs, 20454813386 residues

Word size : 0  
Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb.ba.\*
- 2: gb.htg.\*
- 3: gb.in.\*
- 4: gb.om.\*
- 5: gb.ov.\*
- 6: gb.pat.\*
- 7: gb.ph.\*
- 8: gb.pl.\*
- 9: gb.pr.\*
- 10: gb.ro.\*
- 11: gb.sts.\*
- 12: gb.sy.\*
- 13: gb.un.\*
- 14: gb.vi.\*
- 15: em.ba.\*
- 16: em.fun.\*
- 17: em.hum.\*
- 18: em.in.\*
- 19: em.mu.\*
- 20: em.om.\*
- 21: em.or.\*
- 22: em.ov.\*
- 23: em.pat.\*
- 24: em.ph.\*
- 25: em.pl.\*
- 26: em.ro.\*
- 27: em.sts.\*
- 28: em.un.\*
- 29: em.vi.\*
- 30: em.htg\_hum.\*
- 31: em.htg\_inv.\*
- 32: em.htg\_other.\*
- 33: em.htg\_mus.\*
- 34: em.htg\_pln.\*
- 35: em.htg\_rtd.\*
- 36: em.htg\_mam.\*
- 37: em.htg\_vrt.\*
- 38: em.sy.\*
- 39: em.htgo\_hum.\*
- 40: em.htgo\_mus.\*
- 41: em.htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
C 1	1449	99.9	24000	1	AE004856	AE004856 Pseudomon
2	1248	86.1	5280	1	AY008276	AY008276 Pseudomon
3	44	3.0	301995	1	AY016779	AY016779 Pseudomon
4	24	1.7	1317	6	AX189058	AX189058 Sequence
5	24	1.7	1629	1	ECMURD	X17609 Escherichia
6	24	1.7	2608	1	ECMUROY	X51584 Escherichia
7	24	1.7	11441	1	AE015046	AE015046 Shigella
8	24	1.7	12791	1	AE005185	AE005185 Escherich
9	24	1.7	21757	1	AE000118	AE000118 Escherich
10	24	1.7	28277	6	AX191720	AX191720 Sequence
11	24	1.7	28277	15	EC2MIN	X55034 E. coli 2 m
12	24	1.7	111408	1	EC0110K	D10483 Escherichia
13	24	1.7	248308	2	AC073937	AC073937 Mus muscu
14	24	1.7	259498	2	AC020876	AC020876 Mus muscu
15	24	1.7	281530	1	AP002550	AP002550 Escherich
16	24	1.7	290029	1	AE016978	AE016978 Shigella
17	24	1.7	300409	1	AE016755	AE016755 Escherich
18	24	1.7	329709	1	AP002997	AP002997 Mesorhizo
19	23	1.6	713	9	HSA328412	AJ28412 Homo sapi
20	23	1.6	8599	1	AB004563	AB004563 Terrabact
21	23	1.6	110000	2	AC141230_2	Continuation (3 of
22	23	1.6	311600	1	AE016871	AE016871 Pseudomon
23	24	1.5	31495	1	BUJ33883	U33883 Bradyrhizob
24	22	1.5	207050	1	AL646063	AL646063 Ralstonia
25	22	1.5	300900	1	AP005939	AP005939 Bradyrhiz
26	21	1.4	9077	1	AB032799	AB032799 Chromobac
27	21	1.4	10094	1	AF172851	AF172851 Chromobac
28	21	1.4	11985	1	PPY19106	PPY19106 Pseudomonas
29	21	1.4	13376	1	U72354	U72354 Pseudomonas
30	21	1.4	69301	1	STU82965	STU82965 Streptococ
31	21	1.4	86896	1	RCU57682	RCU57682 Rhodobacte
32	21	1.4	132469	2	AP003926	AP003926 Oryza sat
33	21	1.4	152460	8	AP004617	AP004617 Oryza sat
34	21	1.4	186217	10	AC122310	AC122310 Mus muscu
35	21	1.4	310550	1	SC0939113	SC0939113 Streptoc
36	20	1.4	2076	6	AX196006	AX196006 Sequence
37	20	1.4	3511	9	HSSKIR	X15218 Human ski o
38	20	1.4	3772	1	MELOSVM12	AF049243 Mesorhizo
39	20	1.4	4583	1	SMASFOABC	M33815 S.marcescen
40	20	1.4	7313	1	AF183959	AF183959 Pseudomon
41	20	1.4	10029	1	AE012267	AE012267 Xanthomon
42	20	1.4	10684	1	AE009365	AE009365 Agrobacte
43	20	1.4	10801	1	AE012280	AE012280 Xanthomon
44	20	1.4	12036	1	AE008249	AE008249 Agrobacte
45	20	1.4	12296	1	AF031417	AF031417 Pseudomon

ALIGNMENTS

RESULT 1  
AE004856/c  
LOCUS  
DEFINITION Pseudomonas aeruginosa PAO1, section 417 of the complete genome.  
ACCESSION AE004856 AE004091  
VERSION AE004856.1 GI:9950633  
KEYWORDS  
SOURCE Pseudomonas aeruginosa PAO1  
ORGANISM Pseudomonas aeruginosa PAO1  
Bacteria: Proteobacteria: Gammaproteobacteria; Pseudomonadales;  
Pseudomonadaceae; Pseudomonas.  
REFERENCE 1 (bases 1 to 24000)  
AUTHORS Stover,C.K., Pham,X.-Q.T., Erwin,A.L., Mizoguchi,S.D., Warrenner,P., Hickey,M.J., Brinkman,F.S.L., Hufnagle,W.O., Kowalik,D.J.,

Lagrou, M., Garber, R.L., Goltry, L., Tolentino, E.,  
 Westbrock-Wadman, S., Yuan, Y., Brody, L.L., Coulter, S.N.,  
 Folger, K.R., Kas, A., Larbig, K., Lim, R.M., Smith, K.A., Spencer, D.H.,  
 Wong, G.K.-S., Wu, Z., and Paulsen, I.T.  
 Complete genome sequence of *Pseudomonas aeruginosa* PA01, an  
 opportunistic pathogen  
 Nature 406 (6799), 959-964 (2000)  
 20437337  
 10984043  
 2 (bases 1 to 24000)  
 Stover, C.K., Pham, X.-Q.T., Erwin, A.L., Mizoguchi, S.D., Warren, P.,  
 Hickey, M.J., Brinkman, F.S.L., Huftnagle, W.O., Kowalik, D.J.,  
 Lagrou, M., Garber, R.L., Goltry, L., Tolentino, E.,  
 Westbrock-Wadman, S., Yuan, Y., Brody, L.L., Coulter, S.N.,  
 Folger, K.R., Kas, A., Larbig, K., Lim, R.M., Smith, K.A., Spencer, D.H.,  
 Wong, G.K.-S., Wu, Z., Paulsen, I.T., Reizer, J., Saiter, M.H.,  
 Hancock, R.E.W., Lory, S. and Olson, M.V.  
 Direct Submission  
 Submitted (16-MAY-2000) Department of Medicine and Genetics,  
 University of Washington Genome Center, University Of Washington,  
 Box 352145, Seattle, WA 98195, USA  
 3 (bases 1 to 24000)

**TITLE**  
**JOURNAL**  
**MEDLINE**  
**PUBMED**  
**REFERENCE**  
**AUTHORS**  
**CONSTRM**  
**TITLE**  
**JOURNAL**  
**COMMENT**

*Pseudomonas aeruginosa* Community Annotation Project (PseudocAP)  
 Direct Submission  
 Submitted (04-FEB-2003) Department of Molecular Biology and  
 Biochemistry, Simon Fraser University, 8888 University Dr.,  
 Burnaby, British Columbia V5A 1S6, Canada  
 -----  
 This represents the February 3, 2003 version of the continually  
 updated, reviewed, *Pseudomonas aeruginosa* PA01 genome annotation,  
 from PseudocAP (see <http://www.pseudomonas.com> for latest updates  
 and links to alternate annotations). PseudocAP is coordinated by  
 Fiona S.L. Brinkman (Simon Fraser University, Canada) and Robert  
 E.W. Hancock (University of British Columbia, Canada). We welcome  
 submission through [www.pseudomonas.com](http://www.pseudomonas.com) of any proposed changes.

'Protein name confidence' is used to rate our confidence of the  
 accuracy of the protein name.  
 Class 1: Function experimentally demonstrated in *P. aeruginosa*.  
 Class 2: Function of highly similar gene experimentally  
 demonstrated in another organism (and gene context consistent  
 in terms of pathways its involved in, if known).  
 Class 3: Function proposed based on presence of conserved amino  
 acid motif, structural feature or limited sequence similarity  
 to an experimentally studied gene.  
 Class 4: Homologs of previously reported genes of unknown function,  
 or no similarity to any previously reported sequences.

**FEATURES**  
 source  
 Location/Qualifiers  
 1. .24000  
 /organism="Pseudomonas aeruginosa PA01"  
 /mol\_type="genomic DNA"  
 /strain="PA01"  
 /db\_xref="taxon:208964"  
 complement(91..1308)  
 /gene="argJ"  
 /locus\_tag="PA4402"  
 complement(91..1308)  
 /gene="argJ"  
 /locus\_tag="PA4402"  
 /note="Protein name confidence: Class 2 (high similarity  
 to functionally studied protein); Subcellular  
 localization: Cytoplasmic; Subcellular localization  
 confidence: Class 2; Alt protein name: ornithine  
 transacetylase"  
 /codon\_start=1  
 /transl\_table=11  
 /product="glutamate N-acetyltransferase"  
 /protein\_id="AAG07790.1"  
 /db\_xref="GI:9950634"  
 /translation="MAVGLGPLSLHPVPGFELGIASGKIPKRGKDVVVMRCAGSS  
 VAGVFTLNAFCAPVTLAKQRFELGEVRYLLTWTGNAGTGEAGLAAATCAKLAEL

AGVAETSVLPYSTGVIGEPUPVAKIEAALPAALADLAEDRWAEAAAAGIMTTDTLPKA  
 SROFVDGTVTVTIGISKAGMIKPNMTMLGVIATDAKVAQGVLODLLRDAANKSFN  
 RITIDGDTNDCMLIATGRAALPEVTOAGSALFAALQAVLEVSMELAQAIVRGE  
 GATKFTVQVNGGATHQECLDVGVAVAHSPLIKLTALFASPDPNNGRIILAAVGRAGVANL  
 DVSKIDVFLGDVCIASRGGRAASTTEEGGAAMVAQAEIGIRIELRGTCSETITWTDL  
 SHEYKVINAEYRT"  
 complement(1455..4205)  
 /gene="seca"  
 /locus\_tag="PA4403"  
 complement(1455..4205)  
 /gene="seca"  
 /locus\_tag="PA4403"  
 /note="Protein name confidence: Class 2 (high similarity  
 to functionally studied protein)"  
 /codon\_start=1  
 /transl\_table=11  
 /product="secretion protein SecA"  
 /protein\_id="AAG07791.1"  
 /db\_xref="GI:9950635"  
 /translation="MFAPILLKFLSGKNRDKRMKAKVQAINALEPQMVALSDEQLK  
 AKTAEQRYAKGETLDQLPEAFVAVVYVNDYLARDANMRPLVFLGSLVGVTPF  
 RTGEGTLVGLTPVYLNALSGKGVHVTVNDYLARDANMRPLVFLGSLVGVTPF  
 QPPEDKRAAYAADITGTNNEFGFDYLRDNMAFSLDDKFORELNFAVDEVDLIDE  
 ARTPLISGAQEDSSELIYKINKLIPLAKQVEVEGKPTTEGHISIDERTQVELNE  
 QGHQFIEDLSQNLGEGESLYSAHNSLLTHVYAAALRAHTLHFRNVEYVIGDQIL  
 LIDHTGRTPMGRRLSEGLHQAEKEGLEIOAESQTLASTFYNTFRLYKRLGATGT  
 TADTEAFERQIYGLDVVTPTRPTARKDFNDLVLTQEEKYAAITDITKQCAALGR  
 PILVGASTESSEYVSKLEAGIEHKVLNAKYEKEAIIAQAGAPSVTIATNMAG  
 RGTDIILGGNVEVALENPTEQIAQIKAEWKHQVQVIEAGGLHVIASERHESRR  
 IDNQLGRAGRGQDPCSSRYLSLEDNLRIFASDRVKNFMKALGQSGAIEHRVWT  
 LATEKRAQVGRNDFIRKQLEDDVANEQRKVIYTHMRNLISAEDVGETIEFRE  
 TSATINQHIPQSLPEQDIELEAALYSDFAVRLPQQMLDEDKLYETIREFRE  
 EQVAAVYKEELAGAEALRAFEKMLRLVLDLWDKDLHSLDHLRGIHLRGYAKRN  
 PKQYKRESFTLFQELDSIKRDTIRVLSHVQVRREDPAEEARLRREAEAKRQF  
 QHAEAFSMEQAVAGEEELPEGPVVPVLEPVREQKIGRNEQPCGSGKKYKHCQ  
 LD"  
 complement(4339..5262)  
 /locus\_tag="PA4404"  
 complement(4339..5262)  
 /locus\_tag="PA4404"  
 /note="Protein name confidence: Class 4 (homologs of  
 previously reported genes of unknown function, or no  
 similarity to any previously reported sequences)"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="AAG07792.1"  
 /db\_xref="GI:9950636"  
 /translation="MHIIFLSRHHGAARSVHLQPRLLMLAGACVVGASLLAGFMGAW  
 LFRAPAPAGIEVDAAALQAQSAEISSAREDAQMDALSVHLGELQARLVRLDALGER  
 LTELADVDPEDFSLVSGGGYDEPLEGSAYAPPSPMATLDELAARLREQOLEYL  
 EQLLADKHISQSTYLACQPVQGYISSPFGRRVNPVSGRLSMHKGIDFAPAGSDVYA  
 VAAGVVTSGRKNGYGVYEVGHADGYTTLTAHNQKNLVEVDLVKRGQVLAKVSGSTG  
 RSTGYHVHFEVMDGRVMPNPQSFARASVSE"  
 5409..5804  
 /locus\_tag="PA4405"  
 5409..5804  
 /locus\_tag="PA4405"  
 /note="Protein name confidence: Class 4 (homologs of  
 previously reported genes of unknown function, or no  
 similarity to any previously reported sequences)"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="AAG07793.1"  
 /db\_xref="GI:9950637"  
 /translation="MLFAEAQRICRLQQLLESLOQAAPRYCHVASHRGCLLIIVTD  
 GHWATRLYQRRRLQQLQAEFEFANLAKILFKVQPTTSTQRAARSLTUSAAENL  
 RESAEIVGNPRLREALERLASRAEKRSQE"  
 complement(5865..6776)  
 /gene="lpxC"  
 /locus\_tag="PA4406"  
 /note="synonyms: envA, asmB"

[illegible]

```

RESULT 2
AY008276      5280 bp      DNA      linear      BCT 19-JUL-2001
LOCUS
DEFINITION
Pseudomonas aeruginosa MurE (murE), MurF (murF), MurY (murY), and
MurD (murD) genes, complete cds.
ACCESSION
AY008276
KEYWORDS
AY008276.2 GI:12007715
SOURCE
Pseudomonas aeruginosa
Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
1 (bases 1 to 5280)
Azzolina,B.A., Yuan,X., Anderson,M.S. and El-Sherbeini,M.
The cell wall and cell division gene cluster in the Mra operon of
Pseudomonas aeruginosa: cloning, production, and purification of
active enzymes
PROTEIN
Protein Expr. Purif. 21 (3), 393-400 (2001)
21178826
11287173
2 (bases 1 to 5280)
El-Sherbeini,M. and Azzolina,B.
Pseudomonas aeruginosa mray, murD, murE, and murF genes
Unpublished
3 (bases 1 to 5280)
El-Sherbeini,M. and Azzolina,B.
Direct Submission
Submitted (20-SEP-2000) Biochemistry, Merck Research Laboratories,
126 E. Lincoln Avenue, Rahway, NJ 07065, USA
4 (bases 1 to 5280)
El-Sherbeini,M. and Azzolina,B.
Direct Submission
Submitted (02-JAN-2001) Biochemistry, Merck Research Laboratories,
126 E. Lincoln Avenue, Rahway, NJ 07065, USA
REMARK
Sequence update by submitter
COMMENT
On Jan 2, 2001 this sequence version replaced gi:10719753.
FEATURES
source
1..5280
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/db_xref="taxon:287"
1..1464
/genes="murE"
1..1464
/genes="murE"
/notes="UDP-N-acetylmuramoyl-L-alanyl-D-glutamate:meso-2,
6-diaminopimelate ligase"
/codon_start=1
/transl_table=11
/product="MurE"
/protein_id="AAG45235.1"
/db_xref="GI:12007716"
/translation="MPMSLSQFPQARDLRLRELTLDHGVPRPDLFTLPVGGHDDG
RAHIAADKATGATAVAYAEAGELPPDPAPLIATKVGAAQLSAVAGRTVGEPSGLD
LIGVTGNTKTSVSQLVAOALDLGERCGIVGTGTFYGALESGRHTTPDPLAVOAT
LATLKAGARAVAMEVSSGLDQGRVAALGFDIATVFTNLSRDLHDYHGSEYAAAKA
KLPAPWDLKCRVRLDDDDGRRRLAAGEEQSELITYSDSSAFLYCKREASPTASPTAR
RPSIAADQGLLRSLGRLNLSLAAVALLGLGSLDGDIILFTLPOLQGPVGRMORL
GGDEVLVYVDYVYHTPDALKEALRPHAAARLLCLFCGGDRDRAGKRPALMAIAER
LADDELIVTDNPTREASAIITADIRKFAAADKVTFLPSRGEAIAHLIASAAVDDVVL
LAGKHEDYQIEDGVHRPFSDTQEAERALAAWEVPHA"
1457..2833
/genes="murF"
/genes="murF"
/notes="recombinant enzyme catalyzed the ATP-dependent
addition of D-alanine-D-alanine to the
UDP-N-acetylmuramyl-L-alanine-D-glutamate-m-Dap precursor"
/codon_start=1
/transl_table=11
/product="MurF"
/protein_id="AAG45236.1"

```

QY	302	TCGCTGTGTACAGCGCGCGGAAAGGGCTGCGCATCTCCGGTGACATCGATCTCTTCGC	1382	AGAGGAGTAGCGTGATGCTG	1402
Db	4180	TCGCTGTGTACAGCGCGCGGAAAGGGCTGCGCATCTCCGGTGACATCGATCTCTTCGC	5260	AGAGGAGTAGCGTGATGCTG	5280
QY	362	CGCGAGGCGAAGGCCCGATCGTCCGATACACGGTTCCAAAGCGAAGACCGTGCAC			
Db	4240	CGCGAGGCGAAGGCCCGATCGTCCGATACACGGTTCCAAAGCGAAGACCGTGCAC			
QY	422	CACCTGTGTGGCGAAATGGCGGTGCGCGGACAAAGCGTGTGCGCGGCGCAACCT			
Db	4300	CACCTGTGTGGCGAAATGGCGGTGCGCGGACAAAGCGTGTGCGCGGCGCAACCT			
QY	482	CGGCAACCCCGCGCTCGACCTGCTGGCGGACACATCGAGCTGTGAGTGTGAGCTGTG			
Db	4360	CGGCAACCCCGCGCTCGACCTGCTGGCGGACACATCGAGCTGTGAGTGTGAGCTGTG			
QY	542	GAGCTTCAGCTGGAAACCTGGGATCGCTCAACGCGAGGTGGGACCGTCTGAACGT			
Db	4420	GAGCTTCAGCTGGAAACCTGGGATCGCTCAACGCGAGGTGGGACCGTCTGAACGT			
QY	602	CAGCGAAGACCATATGGATCGCTACGAGCGCATGCTGACTACCACTGGCGCAAGCAGG			
Db	4480	CAGCGAAGACCATATGGATCGCTACGAGCGCATGCTGACTACCACTGGCGCAAGCAGG			
QY	662	GATCTTCGCGGTGCCCGCAGGTGCTGATCGGCAATCGCGCGATGCCCTGACCCGCGCT			
Db	4540	GATCTTCGCGGTGCCCGCAGGTGCTGATCGGCAATCGCGCGATGCCCTGACCCGCGCT			
QY	722	GATCGCGATACCGTGGCGTGTGCTGCTGGCGCTGAACAGCGCGGACTTCAAGCTTT			
Db	4600	GATCGCGATACCGTGGCGTGTGCTGCTGGCGCTGAACAGCGCGGACTTCAAGCTTT			
QY	782	CGGCTGTATCGAGGAAGCGGCGAGAGTGGCTGGCTTCCAGTTCGACAACTGCTGCC			
Db	4660	CGGCTGTATCGAGGAAGCGGCGAGAGTGGCTGGCTTCCAGTTCGACAACTGCTGCC			
QY	842	GTTGGGCAACTGAAGATCCGTGGCGGCCACAACTATTCCAAAGCGCTGCGCGCTGCC			
Db	4720	GTTGGGCAACTGAAGATCCGTGGCGGCCACAACTATTCCAAAGCGCTGCGCGCTGCC			
QY	902	GTTGGGCAACTGGGTGCGGCTGCCCTGTCAGCCCATGCTCGCGGCGCTGAAGCGGTTTC			
Db	4780	GTTGGGCAACTGGGTGCGGCTGCCCTGTCAGCCCATGCTCGCGGCGCTGAAGCGGTTTC			
QY	962	CGGCTGGCTCATCGCTGCGAGTGGTACCGGAGCGGAGGCGTGAAGTACTACGACGA			
Db	4840	CGGCTGGCTCATCGCTGCGAGTGGTACCGGAGCGGAGGCGTGAAGTACTACGACGA			
QY	1022	TTCCAAGGCCACCAACGTGCGGCGCCCTGCGCGCGATCGAGGGGCTGGTGGCGACAT			
Db	4900	TTCCAAGGCCACCAACGTGCGGCGCCCTGCGCGCGATCGAGGGGCTGGTGGCGACAT			
QY	1082	CGACGGCAAGCTGTGCTGCTCGCGGAGAGCGGCAAGGCGCGCATTTCCATGACCT			
Db	4960	CGACGGCAAGCTGTGCTGCTCGCGGAGAGCGGCAAGGCGCGCATTTCCATGACCT			
QY	1142	GGCGAGCGGTCGCGCGCTTCTGCGGCGGTGTGCTGCTGCGGCTGACGCGCGGCT			
Db	5020	GGCGAGCGGTCGCGCGCTTCTGCGGCGGTGTGCTGCTGCGGCTGACGCGCGGCT			
QY	1202	GATTGCCAGGCACTGGGCAACCGGTCGCTGGTGGCGCTGCGCAACCGCTGGAGGAAGC			
Db	5080	GATTGCCAGGCACTGGGCAACCGGTCGCTGGTGGCGCTGCGCAACCGCTGGAGGAAGC			
QY	1262	AGTCGGAGCGCGCGAGCTGGCGCGGAGAGCGGATGCGGTGTGTCGCCGCCCTG			
Db	5140	AGTCGGAGCGCGCGAGCTGGCGCGGAGAGCGGATGCGGTGTGTCGCCGCCCTG			
QY	1322	CGGAGCTTGGACATGTTCAAGAACTTCGAAAGACCGGACCGCTTTCGCCAAAGCCGT			
Db	5200	CGGAGCTTGGACATGTTCAAGAACTTCGAAAGACCGGACCGCTTTCGCCAAAGCCGT			
<hr/>					
QY	1382	AGAGGAGTAGCGTGATGCTG	1402		
Db	5260	AGAGGAGTAGCGTGATGCTG	5280		
<hr/>					
RESULT 3					
LOCUS	AE016779				
DEFINITION	Pseudomonas putida KT2440 section 6 of 21 of the complete genome.				
ACCESSION	AE016779				
VERSION	AE016779.1				
KEYWORDS	GI:26557023				
SOURCE	Pseudomonas putida KT2440				
ORGANISM	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;				
REFERENCE	1 (bases 1 to 301995)				
AUTHORS	Nelson,K., Paulsen,I., Weinel,C., Dodson,R., Hilbert,H., Fouts,D., Gill,S., Pop,M., Martins Dos Santos,V., Holmes,M., Brinkac,L., Beanan,M., DeBoy,R., Daugherty,S., Kolonay,J., Madupu,R., Nelson,W., White,O., Peterson,J., Khouri,H., Hance,I., Lee,P., Holtzapple,E., Scanlan,D., Tran,K., Moazzaz,A., Utterback,T., Rizzo,M., Lee,K., Kosack,D., Moesti,D., Wedler,H., Lauber,J., Hohseil,J., Straetz,M., Heim,S., Kiewitz,C., Eisen,J., Timmis,K., Dueterhoff,A., Tummier,B. and Fraser,C.				
TITLE	Complete genome sequence and comparative analysis of the metabolically versatile Pseudomonas putida KT2440				
JOURNAL	Environ. Microbiol. 4 (12), 799-808 (2002)				
AUTHORS	2 (bases 1 to 301995)				
	Nelson,K., Paulsen,I., Weinel,C., Dodson,R., Hilbert,H., Fouts,D., Gill,S., Pop,M., Martins Dos Santos,V., Holmes,M., Brinkac,L., Beanan,M., DeBoy,R., Daugherty,S., Kolonay,J., Madupu,R., Nelson,W., White,O., Peterson,J., Khouri,H., Hance,I., Lee,P., Holtzapple,E., Scanlan,D., Tran,K., Moazzaz,A., Utterback,T., Rizzo,M., Lee,K., Kosack,D., Moesti,D., Wedler,H., Lauber,J., Hohseil,J., Straetz,M., Heim,S., Kiewitz,C., Eisen,J., Timmis,K., Dueterhoff,A., Tummier,B. and Fraser,C.				
TITLE	Direct Submission				
JOURNAL	Submitted (05-NOV-2002) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA				
FEATURES	Location/Qualifiers				
source	1..301995				
	/organism="Pseudomonas putida KT2440"				
	/mol_type="genomic DNA"				
	/strain="KT2440"				
	/db_xref="taxon:160488"				
gene	89..712				
	/gene="sspA"				
	/locus_tag="PPI320"				
CDS	89..712				
	/gene="sspA"				
	/locus_tag="PPI320"				
	/note="similar to GP:9864180; identified by sequence similarity; putative"				
	/codon_start=1				
	/transl_table=11				
	/product="stringent starvation protein A"				
	/protein_id="AAN66944.1"				
	/db_xref="GI:24982777"				
	/translation="MGATNRLACYSDPADHYSHRVRLVLAEGVSVOLIDVDPALHLPRIKAEVYPGSVPTLVDRDLALYESTVMEYLEIRYPPHLPMPYIPVARGNSRLMHRIQRDWCALADTVLDRSSEAAARKALRESLTGSPLSFSEFACMSDQSLVDCCLLPILRLPVLGIELPQAKPLLDYMERQFAREFQASLSLSSVERMKL"				
	693..1154				
	/gene="sspB"				
	/locus_tag="PPI321"				
	/note="similar to GB:S50157, GB:X60221, SP:P24539, and PID:509291; identified by sequence similarity; putative"				
	/codon_start=1				
	/transl_table=11				





**COMMENT**

```

/mol_type="genomic DNA"
/db_xref="taxon:562"
/map="2min"
98..102
RBS
CDS
/note="pot. ribosome binding site"
111..1193
/note="orf-y (AA 1-360)"
/codon_start=1
/transl_table=11
/protein_id="CAA35932.1"
/db_xref="GI:42059"
/db_xref="SWISS-PROT:P15876"
/transl_table="MLVLAELHVKYSGNFVSYLTFRAIVLSLLTALFTSIWMGPRM
IAHKLQSFQVNRNDSPESHKSRGPTMGIGMTIAVISVLLWAYPSNPYVMVCVL
VHLQGVGIVGVDDYKRVGSKDTPGLIARHKYFMSVIALGVAFALYLAKGKDTPTQAL
VWPFKDVMPQLGFLYLLAYFVIVGTGNVNLTDGLDGLIIMPTVFAVGGFALVAVAA
TGNNFASYLHPIFLRHAGELVIVCTAIVGAGLGFUWNTFYPAQVFMGDDVGSALGGA
LGIATVLRQFEFLVINGGVFVYVETLSVILQVGSFKLRQIRIFRMAPIHHHYELKMG
EPRIVVRFTIISLMVLVILIGLATLKR"
1018..1023
promoter
/note="pot. -35 region"
1043..1048
promoter
/note="pot. -10 region"
1181..1185
RBS
/note="pot. ribosome binding site"
1196..1212
CDS
/note="murD gene product (AA 1-438)"
/codon_start=1
/transl_table=11
/protein_id="CAA35933.1"
/db_xref="GI:42060"
/db_xref="SWISS-PROT:P14900"
/transl_table="MADYQCKNVYIIGLTLGLSCVDFFLARGVTFRVMDTRMTPPGL
AKPIAVERHTGSLNDEWLMAADLIVASPGIALHSPLSAAADAGIEIVGDIELFCRE
DOAPIVAITGSGKSTVTTLVGEAKAAGVNVGVGNIGLPALMLLDDCELYVLELS
SFOLETSSLOAVAAIILNVTEDHMDRYPFGLQYRAAKLRIENAKYCVVNNADALT
MPINGADERGVSGVNNQDTHLNQOGETVLRVKGKVLNVKEMKLGSGHNYTNALAG
LALADAGLPRASLKALTTFTGLPHRFVIVHNGVRVINDSKATNVGSTEALANG
HVDPTGTHLLGGDGKADSFPLARYLQNDNRLYCFGRDGAQLAALRPEVAQETETME
QAMRLAPRVQGDGMVLLSPACSLDQKFNQEGRNEFARLAKELG"
530 a 631 c 777 g 670 t
BASE COUNT
ORIGIN

Query Match 1.78; Score 24; DB 1; Length 2608;
Best Local Similarity 100.0%; Pred. No. 4.7;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 537 CTGTCGAGCTCCAGCTGGAAC 560
|||||
Db 1670 CTGTCGAGCTCCAGCTGGAAC 1693
|||||

RESULT 7
AE015046
LOCUS Shigella flexneri 2a str. 301 section 9 of 412 of the complete
genome.
ACCESSION AE015046
VERSION AE015046.1 GI:24050283
KEYWORDS
SOURCE Shigella flexneri 2a str. 301
ORGANISM Shigella flexneri 2a str. 301
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Shigella.
1 (bases 1 to 11441)
Jin,Q., Yuan,Z.H., Xu,J.G., Wang,Y., Shen,Y., Lu,W.C., Wang,J.H.,
Liu,H., Yang,J., Yang,F., Qu,D., Zhang,X.B., Zhang,J.Y., Zhang,G.W.,
Wu,H.T., Dong,J., Sun,L.L., Xue,Y., Zhao,A.L., Gao,Y.S., Zhu,J.P.,
Kan,B., Chen,S.X., Yao,Z.J., He,B.K., Chen,R.S., Ma,D.L.,
Qiang,B.Q., Wen,Y.M., Hou,Y.D. and Yu,J.
Genome sequence of Shigella flexneri 2a: insights into
pathogenicity through comparison with genomes of Escherichia coli
K12 and O157

```

gene	EKLQHQHVDPPSQENIVVQK" 2107..3873 /gene="fts1" /locus_tag="SF0081" 2107..3873 /gene="fts1" /locus_tag="SF0081" /note="Residues 1 to 588 of 588 are 99 pct identical to residues 1 to 588 of a 588 aa protein from Escherichia coli O157:H7 EDL933 ref: NP_285780.1" /codon_start=1 /transl_table=11 /product="septum formation; penicillin-binding protein 3; peptidoglycan synthetase" /protein_id="AA041745.1" /db_xref="GI:24050287"
CDS	/translation="MKAARKTKPKRQEHANFTISWRPALLCGCILLALAFELGRVAV LQVTPDLVREGMRSLRQVSTSRGMITDRSGRLAVSPVKAIWADPKVEHDAG GISVGRWALANALNIPDLQSARINANPKRFTYLARQVNSOMADYIKKLKLPGRH LREISRRYPSPGVTYAHIGFTNVDSQIEGVKSFQDLWGTQGPQERIVRDRYGRVI EDISSTDQAHNLAISIDERLQALVRELNAVAFKAESGSAVLVDVNTGEVLAMA NPSYNLNSLSTGSKPEAMRNTITDVEPGSTVAPMVVNTALQGVREVSYLNTIPY RINGHEIKDARYSELITGLVQKSSNVGSKLALTPSSALVDYISFSGLGKATNLG LVGERSGLYPQKRWSDIERATFSFGYGLMVTPLQARVYATIGSYGRPLSTIKVD PPVPGERYFPESIVRTVVHMSVALPGGGVKAIAKGYRIARTKTAKKVPDGRYI NKYIATAGVAPALVPALVVVINDPQAGKYGGAVSPVFGAIMGVLTMTNIEPD ALTGDKNEFVINDQEGTGGRS" 3860..5347 /gene="murF" /locus_tag="SF0082" 3860..5347 /gene="murF" /locus_tag="SF0082" /EC_number="6.3.2.13" /note="Residues 1 to 495 of 495 are 99 pct identical to residues 1 to 495 of a 495 aa protein from Escherichia coli pir: S40595" /codon_start=1 /transl_table=11 /product="UDP-N-acetylmuramoylalanyl-D-glutamate-2,6-diaminopimelate ligase" /protein_id="AA041747.1" /db_xref="GI:24050288"
gene	/translation="MADRNLRLDLPAPVPDAPSRALREMTLDSRVAAAGDLFVAVVGH QADGRYIPQAIQGVAAIIEAKDEATDGEIRMHGVPVIVLSOLNERLSALGRFV HEPSDNLRLVGTGNGKTTTOLLQAWSQLLGETSAVMGTGNGLLGKVIPTENTTG SNVDVQHELGLVDCGATFCAMEVSSHGLVQHRVAALKFPASVFTNLSRDHLHDGDM EHYEAARKLLISEHCGQATINADDEVGRWMLKPLDAPAVSMEDHINPNCGRWLKA TEVNYHDSGATIRFESSWSGDEIEHLMGFNFVNLALATLALGYPLADLLKTA RLQPCGMEVFTAFGKPTVVYDYAHTDALEKALQAAHLCAKGLWCVFCGGGRDK GKRPLMGATIEEFADVAVVTDNPTPEPRATINDILAGMLDAGHAKVMEGRAEAVTC AVMQAKENDVVLVAGKHEDYQIVGNQRLDYSDRVTVVARLLGGIA" 5344..6702 /gene="murF" /locus_tag="SF0083" 5344..6702 /gene="murF" /locus_tag="SF0083" /note="Residues 1 to 452 of 452 are 99 pct identical to residues 1 to 452 of a 452 aa protein from Escherichia coli K12 ref: NP_414628.1" /codon_start=1 /transl_table=11 /product="D-alanine-D-alanine-adding enzyme" /protein_id="AA041748.1" /db_xref="GI:24050289"
CDS	/translation="MISVTLSQLTDILNGELQAGADITLDVTTDFRKLTPOGCLFVALK GERFDHAFDAQAGAGALVSRPLDIDLQILVTKDTRLAFGLAARVQOQPALV VALTGSGKTSVKEMTAAILSOCGNTLVYTAGNLNDIGVPMWTLRLLTPEYDVAIEIG ANHOCEIAWTVSLTPEALVNNLAAHLECGFSLAGVAKAGEIFSGLPENGIAIMN ADNDRLNWSQVIGSRKVRSPNANSDFATNLHVTHSGMEFTLOPTGTSVDVLLP LPGRNINAAALAAALSMSVGTDLAIKAGLANLKAIVPGRLFFPIQIAENQLLLDSDYN ANVGSMTAAVQVLAEMPGYRVLVVGDMALGAESSEACHVQVGEAKAAGIDRVLSMKG
gene	QSHAISTASGVGEHFADKLTALITRLKSLIAEQVQVITILVKGSRSAAMEEVVRLAOENG TC" 5696..7778 /gene="mraI" /locus_tag="SF0084" 5696..7778 /gene="mraI" /locus_tag="SF0084" /note="Residues 1 to 360 of 360 are 100 pct identical to residues 1 to 360 of a 360 aa protein from Escherichia coli K12 ref: NP_414629.1" /codon_start=1 /transl_table=11 /product="phospho-N-acetylmuramoyl-pentapeptide transferase" /protein_id="AA041749.1" /db_xref="GI:24050290"
CDS	/translation="MLVWLAELHVKYSGFNVSFLYTRAFIVSLTALFISLMMGPRM IAHLQKLSFGQVRNDGSPESHFSKRGTPMTGGIMILTAIVISLVLLWAYSPNYPVWCVL VVLGYGVIGFVDDYRKVVKDKTGLIARKWYFWMVSIALGVAFALYLAGDTPATQL VVPFKDVPOLGLFYIILAVFIVGTGNVNLDTGLDGLAIIMPVTFVAGGFALVAMA TGNNFASYLHPIPLRHAGELVIVCTAIVAGLGFELWENTYPAQVFMGDVGSALGGA LGITAVLRQFELLVIMGVFVETLSVILQVSGFKLRGQRIFRMAPIHMHYELAGWP EPRVTVRMIISLMLVLIGLATLKVR" 7781..9097 /gene="murD" /locus_tag="SF0085" 7781..9097 /gene="murD" /locus_tag="SF0085" /note="Residues 1 to 438 of 438 are 99 pct identical to residues 1 to 438 of a 438 aa protein from Escherichia coli K12 ref: NP_414630.1" /codon_start=1 /transl_table=11
gene	Query Match 1.7%; Score 24; DB 1; Length 11441; Best Local Similarity 100.0%; Pred. No. 3.7; Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CDS	QY 537 CTGTCGAGCTTCCAGCTGGAAACC 560 DB 8255 CTGTCGAGCTTCCAGCTGGAAACC 8278 RESULT 8 AE005185 LOCUS 12791 bp DNA linear BCT 21-MAR-2001 DEFINITION Escherichia coli O157:H7 EDL933 genome, contig 1 of 3, section 9 of 155. ACCESSION AE005185 AE005174 VERSION AE005185.1 GI:12512782 KEYWORDS SOURCE Escherichia coli O157:H7 EDL933 ORGANISM Escherichia coli O157:H7 EDL933 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia. REFERENCE 1 (bases 1 to 12791) AUTHORS Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D., Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A., Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L., Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K., Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C., Welch,R.A. and Blattner,F.R. TITLE Genome sequence of enterohaemorrhagic Escherichia coli O157:H7 JOURNAL Nature 409 (6819), 529-533 (2001) MEDLINE 21074935 PUBMED 11206551 REFERENCE 2 (bases 1 to 12791) AUTHORS Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D., Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A., Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L., Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K., Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,

Apodaca, J., Anantharaman, T.S., Lin, J., Yen, G., Schwartz, D.C.,  
 Welch, R.A. and Blattner, F.R.  
 TITLE Direct Submission  
 JOURNAL Submitted (22-OCT-2000) Laboratory of Genetics, University of  
 Wisconsin, 445 Henry Mall, Madison, WI 53706, USA

FEATURES  
 source location/Qualifiers  
 1. .12791  
 /organism="Escherichia coli O157:H7 EDL933"  
 /mol\_type="genomic DNA"  
 /strain="EDL933"  
 /serotype="O157:H7"  
 /db\_xref="taxon:155864"  
 /note="enterohemorrhagic"  
 72. .1076  
 /gene="fru"  
 /note="synonym: Z0090"  
 72. .1076  
 /gene="fru"  
 /function="regulator; Energy metabolism, carbon:  
 Glycolysis"  
 /notes="Residues 1 to 334 of 334 are 100.00 pct identical  
 to residues 1 to 334 of 334 from Escherichia coli K-12  
 Strain MG1655: B0080"  
 /codon\_start=1  
 /transl\_table=11  
 /product="transcriptional repressor of fru operon and  
 others"  
 /protein\_id="AAG54384.1"  
 /db\_xref="GI:12512783"  
 /translation="MKLDEIARLGVSRRTASVINGKAKQYRVSKTEKVMVAVRE  
 HNYHPNVAAGLRAGTRISGLVLPDLENTSYRIANYLERQARQYGLLIACSEDQ  
 PNEMRCIEHLQRQDAIVSTSLPEHPFYQWANDPFPPIVALDRALDRHFTSVV  
 QADODAEMLRELPFAETVLVYLGALPELSVFLREGQFTAKWDDPDEVHFLYAN  
 SYDERAAQLFEKLTHTPQALFTTSEALQGVMDVTLRDKGLPSLATATFGDN  
 ELLDFQCPLVLAQRHRDVAERVLVILASLDLDEPKRPGLTRIARNLYRQVLRS  
 1678. .2136  
 /gene="yabB"  
 /note="synonym: Z0091"  
 1678. .2136  
 /gene="yabB"  
 /function="orf; Unknown function"  
 /note="Residues 1 to 152 of 152 are 99.34 pct identical to  
 residues 1 to 152 of 152 from Escherichia coli K-12 Strain  
 MG1655: B0081"  
 /codon\_start=1  
 /transl\_table=11  
 /product="orf, hypothetical protein"  
 /protein\_id="AAG54385.1"  
 /db\_xref="GI:12512784"  
 /translation="MFRGATLVNLSKGRSLVPTRYRREOLLENAAGQMVCTIDIHPC  
 LLLYPLPEWEIIEOKLSLSSMNPVRRVORLLIGHASEQMDGAGRLLIAPVLRQHA  
 GLTKEWLVGQFNKFLWDETTHQVQKEDIDAEQLATGDLSERLQDLSI"  
 2138. .3079  
 /gene="yabC"  
 /note="synonym: Z0092"  
 2138. .3079  
 /gene="yabC"  
 /function="orf; Not classified"  
 /note="Residues 1 to 313 of 313 are 100.00 pct identical  
 to residues 1 to 313 of 313 from Escherichia coli K-12  
 Strain MG1655: B0082"  
 /codon\_start=1  
 /transl\_table=11  
 /product="putative apolipoprotein"  
 /protein\_id="AAG54386.1"  
 /db\_xref="GI:12512785"  
 /translation="WMENYKHTVTLLDEAVNGLINRPDGIYIDGTFGRGGHRLILSQ  
 LGEGRLLADRDQPAIAVAKTIDDPFSIIHGPFSAIGEYVAERDLKIDGILDL  
 GVSQDLDAERGFSGFMRDQPLMDRPGQSAEAEWLQTAEEADTAWLKYGERF  
 AKRIARIVERNRQOPMTKELAEVVAATPVKDKFHPATRTFOAVRIWNSLEE  
 IEQALKSSILNVLPGRLSIIISFSLSEDRIVKFRNRSRGFPQVAGLPMTQEQLKL  
 GGRQLRALGKLMFGEEVAENPRAKSSVLRIARTNA"

gene  
 3076. .3441  
 /gene="ftsL"  
 /note="synonym: Z0093"  
 3076. .3441  
 /gene="ftsL"  
 /function="phenotype; Cell division"  
 /note="Residues 1 to 121 of 121 are 100.00 pct identical  
 to residues 1 to 121 of 121 from Escherichia coli K-12  
 Strain MG1655: B0083"  
 /codon\_start=1  
 /transl\_table=11  
 /product="cell division protein; ingrowth of wall at  
 septum"  
 /protein\_id="AAG54387.1"  
 /db\_xref="GI:12512786"  
 /translation="MISRYTEALSKVSGMSHERHALPCVIGDDLLREGLPLCLFI  
 CILTAVVVTTHAHTRLRLTAQREQLVLERDALDIEWRNLIILENALGDHSRVERIAT  
 EKLQMQRVDPQENIVVQK"  
 3457. .5223  
 /gene="ftsI"  
 /note="synonym: Z0094"  
 3457. .5223  
 /gene="ftsI"  
 /function="enzyme; Cell division"  
 /note="Residues 1 to 588 of 588 are 100.00 pct identical  
 to residues 1 to 588 of 588 from Escherichia coli K-12  
 Strain MG1655: B0084"  
 /codon\_start=1  
 /transl\_table=11  
 /product="septum formation; penicillin-binding protein 3;  
 peptidoglycan synthetase"  
 /protein\_id="AAG54388.1"  
 /db\_xref="GI:12512787"  
 /translation="MKAATQKPKQKEEHANFISHRFALLCCILLALAFLLGRVAM  
 LQVSPMLVKEGDMRSLRVQVTSRGMITDSRPLAVSPVPKAIADWPKRVHONG  
 GISVGRKALANALNPLDQLSARINANPKGFRFYLAQVNPDMADYIKKLLPGIHL  
 LREESRYYPGSEVTAHLIGFTNVDQSGIEGKSFDKWLTGQGERIVKRDYGRVI  
 EDISSTDQAANLALSIDERLQALVYRELNNVAEKAEESGSAVLVDVNTGVLAMA  
 NSPSNPNNLSGTPKEAMRNTITDVFEPGSTVKPMVMTALQGVVRENSVLNTIPY  
 RINGHEIKDVARYSLETLTGVLQKSSNVGSKLAMPSSALVDVTSRGLGKATNIG  
 LVNGRSLYPQKQWSDIERAFESFGYGLMVTPLQARVATIGSGIYRPLSITKVD  
 PVPGRVFPPEISIVTVVHMESVALPGGGGVKAAIKGYRIATKTKRVGPDGRIY  
 NIYIATAGVAPASQPRFALVVVINDPQAGKYGGAVSAPVFGAINGGVLRTMNIEPD  
 ALLTGDKNEFVINGQSGTGGRS"  
 5210. .6697  
 /gene="mure"  
 /note="synonym: Z0095"  
 5210. .6697  
 /gene="mure"  
 /function="enzyme; Cell envelop: Murein sacculus,  
 peptidoglycan"  
 /note="Residues 1 to 495 of 495 are 98.98 pct identical to  
 residues 1 to 495 of 495 from Escherichia coli K-12 Strain  
 MG1655: B0085"  
 /codon\_start=1  
 /transl\_table=11  
 /product="meso-diaminopimelate-adding enzyme"  
 /protein\_id="AAG54389.1"  
 /db\_xref="GI:12512788"  
 /translation="MADRNLRLDLPWPDPASRALREMTLDSRVAAGDLFVAVVGH  
 QADGRYIPQAIQAQVAAIIAEAKGEATDGEIRMHGVPVIYLSQLNERLSALAGRFY  
 HEPSONLXXVGTGTNGTKTTTQLLAQSQLIGETSAVGTGNGLLGRVITENTTG  
 SAVDYOHELAVLQDQATFCAMEVSHGLVQHRVAALKFAASVFTNLSRDHLDYHGM  
 EHYEAKWLLYSEHCGQAIINADDEVRRLAKLPDAVAVSMEDHINPNCHRWLKA  
 TEVNVHDSGATIRFSSMGDETESHMGAFVSNLLLATLALLGYPLADLLKTA  
 RLQPCGRMEVTFPGKPVVVDYHTPDALAKQAARLHCAKGLKWCVCVGGCGDRDK  
 GKRPLKAEEDFADVAVVYTDNPTPEPRALINDILAGMLDAGHAKVMEDRAEAVTC  
 AYMKAKENIVLVAGKHEDYQIVGNQRDLSDRVTVARLLGVIA"  
 6694. .8052  
 /gene="murf"  
 /note="synonym: Z0096"  
 6694. .8052  
 /gene="murf"

gene  
 3076. .3441  
 /gene="ftsL"  
 /note="synonym: Z0093"  
 3076. .3441  
 /gene="ftsL"  
 /function="phenotype; Cell division"  
 /note="Residues 1 to 121 of 121 are 100.00 pct identical  
 to residues 1 to 121 of 121 from Escherichia coli K-12  
 Strain MG1655: B0083"  
 /codon\_start=1  
 /transl\_table=11  
 /product="cell division protein; ingrowth of wall at  
 septum"  
 /protein\_id="AAG54387.1"  
 /db\_xref="GI:12512786"  
 /translation="MISRYTEALSKVSGMSHERHALPCVIGDDLLREGLPLCLFI  
 CILTAVVVTTHAHTRLRLTAQREQLVLERDALDIEWRNLIILENALGDHSRVERIAT  
 EKLQMQRVDPQENIVVQK"  
 3457. .5223  
 /gene="ftsI"  
 /note="synonym: Z0094"  
 3457. .5223  
 /gene="ftsI"  
 /function="enzyme; Cell division"  
 /note="Residues 1 to 588 of 588 are 100.00 pct identical  
 to residues 1 to 588 of 588 from Escherichia coli K-12  
 Strain MG1655: B0084"  
 /codon\_start=1  
 /transl\_table=11  
 /product="septum formation; penicillin-binding protein 3;  
 peptidoglycan synthetase"  
 /protein\_id="AAG54388.1"  
 /db\_xref="GI:12512787"  
 /translation="MKAATQKPKQKEEHANFISHRFALLCCILLALAFLLGRVAM  
 LQVSPMLVKEGDMRSLRVQVTSRGMITDSRPLAVSPVPKAIADWPKRVHONG  
 GISVGRKALANALNPLDQLSARINANPKGFRFYLAQVNPDMADYIKKLLPGIHL  
 LREESRYYPGSEVTAHLIGFTNVDQSGIEGKSFDKWLTGQGERIVKRDYGRVI  
 EDISSTDQAANLALSIDERLQALVYRELNNVAEKAEESGSAVLVDVNTGVLAMA  
 NSPSNPNNLSGTPKEAMRNTITDVFEPGSTVKPMVMTALQGVVRENSVLNTIPY  
 RINGHEIKDVARYSLETLTGVLQKSSNVGSKLAMPSSALVDVTSRGLGKATNIG  
 LVNGRSLYPQKQWSDIERAFESFGYGLMVTPLQARVATIGSGIYRPLSITKVD  
 PVPGRVFPPEISIVTVVHMESVALPGGGGVKAAIKGYRIATKTKRVGPDGRIY  
 NIYIATAGVAPASQPRFALVVVINDPQAGKYGGAVSAPVFGAINGGVLRTMNIEPD  
 ALLTGDKNEFVINGQSGTGGRS"  
 5210. .6697  
 /gene="mure"  
 /note="synonym: Z0095"  
 5210. .6697  
 /gene="mure"  
 /function="enzyme; Cell envelop: Murein sacculus,  
 peptidoglycan"  
 /note="Residues 1 to 495 of 495 are 98.98 pct identical to  
 residues 1 to 495 of 495 from Escherichia coli K-12 Strain  
 MG1655: B0085"  
 /codon\_start=1  
 /transl\_table=11  
 /product="meso-diaminopimelate-adding enzyme"  
 /protein\_id="AAG54389.1"  
 /db\_xref="GI:12512788"  
 /translation="MADRNLRLDLPWPDPASRALREMTLDSRVAAGDLFVAVVGH  
 QADGRYIPQAIQAQVAAIIAEAKGEATDGEIRMHGVPVIYLSQLNERLSALAGRFY  
 HEPSONLXXVGTGTNGTKTTTQLLAQSQLIGETSAVGTGNGLLGRVITENTTG  
 SAVDYOHELAVLQDQATFCAMEVSHGLVQHRVAALKFAASVFTNLSRDHLDYHGM  
 EHYEAKWLLYSEHCGQAIINADDEVRRLAKLPDAVAVSMEDHINPNCHRWLKA  
 TEVNVHDSGATIRFSSMGDETESHMGAFVSNLLLATLALLGYPLADLLKTA  
 RLQPCGRMEVTFPGKPVVVDYHTPDALAKQAARLHCAKGLKWCVCVGGCGDRDK  
 GKRPLKAEEDFADVAVVYTDNPTPEPRALINDILAGMLDAGHAKVMEDRAEAVTC  
 AYMKAKENIVLVAGKHEDYQIVGNQRDLSDRVTVARLLGVIA"  
 6694. .8052  
 /gene="murf"  
 /note="synonym: Z0096"  
 6694. .8052  
 /gene="murf"

gene  
 3076. .3441  
 /gene="ftsL"  
 /note="synonym: Z0093"  
 3076. .3441  
 /gene="ftsL"  
 /function="phenotype; Cell division"  
 /note="Residues 1 to 121 of 121 are 100.00 pct identical  
 to residues 1 to 121 of 121 from Escherichia coli K-12  
 Strain MG1655: B0083"  
 /codon\_start=1  
 /transl\_table=11  
 /product="cell division protein; ingrowth of wall at  
 septum"  
 /protein\_id="AAG54387.1"  
 /db\_xref="GI:12512786"  
 /translation="MISRYTEALSKVSGMSHERHALPCVIGDDLLREGLPLCLFI  
 CILTAVVVTTHAHTRLRLTAQREQLVLERDALDIEWRNLIILENALGDHSRVERIAT  
 EKLQMQRVDPQENIVVQK"  
 3457. .5223  
 /gene="ftsI"  
 /note="synonym: Z0094"  
 3457. .5223  
 /gene="ftsI"  
 /function="enzyme; Cell division"  
 /note="Residues 1 to 588 of 588 are 100.00 pct identical  
 to residues 1 to 588 of 588 from Escherichia coli K-12  
 Strain MG1655: B0084"  
 /codon\_start=1  
 /transl\_table=11  
 /product="septum formation; penicillin-binding protein 3;  
 peptidoglycan synthetase"  
 /protein\_id="AAG54388.1"  
 /db\_xref="GI:12512787"  
 /translation="MKAATQKPKQKEEHANFISHRFALLCCILLALAFLLGRVAM  
 LQVSPMLVKEGDMRSLRVQVTSRGMITDSRPLAVSPVPKAIADWPKRVHONG  
 GISVGRKALANALNPLDQLSARINANPKGFRFYLAQVNPDMADYIKKLLPGIHL  
 LREESRYYPGSEVTAHLIGFTNVDQSGIEGKSFDKWLTGQGERIVKRDYGRVI  
 EDISSTDQAANLALSIDERLQALVYRELNNVAEKAEESGSAVLVDVNTGVLAMA  
 NSPSNPNNLSGTPKEAMRNTITDVFEPGSTVKPMVMTALQGVVRENSVLNTIPY  
 RINGHEIKDVARYSLETLTGVLQKSSNVGSKLAMPSSALVDVTSRGLGKATNIG  
 LVNGRSLYPQKQWSDIERAFESFGYGLMVTPLQARVATIGSGIYRPLSITKVD  
 PVPGRVFPPEISIVTVVHMESVALPGGGGVKAAIKGYRIATKTKRVGPDGRIY  
 NIYIATAGVAPASQPRFALVVVINDPQAGKYGGAVSAPVFGAINGGVLRTMNIEPD  
 ALLTGDKNEFVINGQSGTGGRS"  
 5210. .6697  
 /gene="mure"  
 /note="synonym: Z0095"  
 5210. .6697  
 /gene="mure"  
 /function="enzyme; Cell envelop: Murein sacculus,  
 peptidoglycan"  
 /note="Residues 1 to 495 of 495 are 98.98 pct identical to  
 residues 1 to 495 of 495 from Escherichia coli K-12 Strain  
 MG1655: B0085"  
 /codon\_start=1  
 /transl\_table=11  
 /product="meso-diaminopimelate-adding enzyme"  
 /protein\_id="AAG54389.1"  
 /db\_xref="GI:12512788"  
 /translation="MADRNLRLDLPWPDPASRALREMTLDSRVAAGDLFVAVVGH  
 QADGRYIPQAIQAQVAAIIAEAKGEATDGEIRMHGVPVIYLSQLNERLSALAGRFY  
 HEPSONLXXVGTGTNGTKTTTQLLAQSQLIGETSAVGTGNGLLGRVITENTTG  
 SAVDYOHELAVLQDQATFCAMEVSHGLVQHRVAALKFAASVFTNLSRDHLDYHGM  
 EHYEAKWLLYSEHCGQAIINADDEVRRLAKLPDAVAVSMEDHINPNCHRWLKA  
 TEVNVHDSGATIRFSSMGDETESHMGAFVSNLLLATLALLGYPLADLLKTA  
 RLQPCGRMEVTFPGKPVVVDYHTPDALAKQAARLHCAKGLKWCVCVGGCGDRDK  
 GKRPLKAEEDFADVAVVYTDNPTPEPRALINDILAGMLDAGHAKVMEDRAEAVTC  
 AYMKAKENIVLVAGKHEDYQIVGNQRDLSDRVTVARLLGVIA"  
 6694. .8052  
 /gene="murf"  
 /note="synonym: Z0096"  
 6694. .8052  
 /gene="murf"



```

/note="central position to predicted promoter: -319.5"
/round_moiety="fyrR predicted site"
554. .565
/note="central position to predicted promoter:64"
/round_moiety="AraC predicted site"
646. .1767
/genes="leu0"
/note="synonym: b0076"
646. .1767
/genes="leu0"
/note="central position to predicted promoter:156.5"
/round_moiety="Lrp predicted site"
2017. .2045
/genes="ilvI"
/note="factor Sigma70; promoter ilvIHp1; documented+1 at
85597"
3812. .4303
/genes="ilvH"
/note="synonym: b0078"
3812. .4303
/genes="ilvH"
/EC_number="4.1.3.18"
/protein_bind
promoter
gene
CDS

Query Match 1.7%; Score 24; DB 1; Length 21757;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps -0;

QY 537 CTGTCGAGCTTCAGCTGGAACCC 560
|||||
Db 14016 CTGTCGAGCTTCAGCTGGAACCC 14039

RESULT 10
AX191720
LOCUS AX191720 28277 bp DNA linear PAT 15-AUG-2001
DEFINITION Sequence 2 from Patent WO0149775.
ACCESSION AX191720
VERSION AX191720.1 GI:15209889
KEYWORDS
SOURCE Escherichia coli
ORGANISM Escherichia coli
REFERENCE 1
AUTHORS Iversen,P.L.
TITLE Antisense antibacterial cell division composition and method
JOURNAL Patent: WO 0149775-A 2 12-JUL-2001;
Avi Biopharma, Inc. (US)
FEATURES
source
1. .28277
/organism="Escherichia coli"
/mol_type="genomic DNA"
/db_xref="taxon:562"
BASE COUNT 6714 a 6725 c 7993 g 6845 t
ORIGIN

Query Match 1.7%; Score 24; DB 6; Length 28277;
Best Local Similarity 100.0%; Pred. No. 3.2;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 537 CTGTCGAGCTTCAGCTGGAACCC 560
|||||
Db 14091 CTGTCGAGCTTCAGCTGGAACCC 14114

RESULT 11
EC2MIN standard; DNA; PRO: 28277 BP.
ID EC2MIN
XX
XX X55034; M10429;
XX AC
XX SV
XX X55034.1
XX
XX 21-FEB-1991 (Rel. 27, Created)

```

```

/note="central position to predicted promoter: -319.5"
/round_moiety="fyrR predicted site"
554. .565
/note="central position to predicted promoter:64"
/round_moiety="AraC predicted site"
646. .1767
/genes="leu0"
/note="synonym: b0076"
646. .1767
/genes="leu0"
/note="central position to predicted promoter:156.5"
/round_moiety="Lrp predicted site"
2017. .2045
/genes="ilvI"
/note="factor Sigma70; promoter ilvIHp1; documented+1 at
85597"
3812. .4303
/genes="ilvH"
/note="synonym: b0078"
3812. .4303
/genes="ilvH"
/EC_number="4.1.3.18"
/protein_bind
promoter
gene
CDS

Query Match 1.7%; Score 24; DB 1; Length 21757;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps -0;

QY 537 CTGTCGAGCTTCAGCTGGAACCC 560
|||||
Db 14016 CTGTCGAGCTTCAGCTGGAACCC 14039

RESULT 10
AX191720
LOCUS AX191720 28277 bp DNA linear PAT 15-AUG-2001
DEFINITION Sequence 2 from Patent WO0149775.
ACCESSION AX191720
VERSION AX191720.1 GI:15209889
KEYWORDS
SOURCE Escherichia coli
ORGANISM Escherichia coli
REFERENCE 1
AUTHORS Iversen,P.L.
TITLE Antisense antibacterial cell division composition and method
JOURNAL Patent: WO 0149775-A 2 12-JUL-2001;
Avi Biopharma, Inc. (US)
FEATURES
source
1. .28277
/organism="Escherichia coli"
/mol_type="genomic DNA"
/db_xref="taxon:562"
BASE COUNT 6714 a 6725 c 7993 g 6845 t
ORIGIN

Query Match 1.7%; Score 24; DB 6; Length 28277;
Best Local Similarity 100.0%; Pred. No. 3.2;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 537 CTGTCGAGCTTCAGCTGGAACCC 560
|||||
Db 14091 CTGTCGAGCTTCAGCTGGAACCC 14114

RESULT 11
EC2MIN standard; DNA; PRO: 28277 BP.
ID EC2MIN
XX
XX X55034; M10429;
XX AC
XX SV
XX X55034.1
XX
XX 21-FEB-1991 (Rel. 27, Created)

```

DT 05-JUL-1999 (Rel. 60, Last updated, Version 31)  
XX E. coli 2 minute region  
KW ddl gene; envA gene; ftsA gene; ftsQ gene; ftsW gene; ftsZ gene; ilvH gene;  
KW ilvI gene; leuA gene; leuO gene; mraY gene; murC gene; murD gene;  
KW murF gene; murG gene; mutT gene; orfA; orfB; orfC; orfX;  
KW pbbB gene; secA gene; shi gene.  
XX  
OS Escherichia coli  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
XX  
RN [1]  
RP 1-28277  
RA Ayala J.A.;  
RT  
RL Submitted (08-JAN-1991) to the EMBL/GenBank/DBJ databases.  
RL Ayala J.A., Instituto de Biología Molecular, Centro de Biología Molecular,  
RL Universidad Autónoma, Canto-Bianco 28049, Madrid, Spain.  
XX  
RN [2]  
RA Ayala J.A.;  
RT "Regulation of transcription at the 2-minute region of the genetic map of  
RT Escherichia coli";  
RT Unpublished.  
XX  
RN [3]  
RP 1-306  
RX MEDLINE; 82078077.  
RX PUBMED; 6171647.  
RX Wessler S.R., Calvo J.M.;  
RA "Control of leu operon expression in Escherichia coli by a transcription  
RT attenuation mechanism";  
RL J. Mol. Biol. 149(4):579-597(1981).  
XX  
RN [4]  
RP 268-1130  
RX MEDLINE; 86223773.  
RX PUBMED; 3519576.  
RX Haughn G.W., Wessler S.R., Gemmill R.M., Calvo J.M.;  
RA "High A + T content conserved in DNA sequences upstream of leuABCD in  
RT Escherichia coli and Salmonella typhimurium";  
RL J. Bacteriol. 166(3):1113-1117(1986).  
XX  
RN [5]  
RP 843-1812  
RX MEDLINE; 88320486.  
RX PUBMED; 3413113.  
RA Henikoff S., Haughn G.W., Calvo J.M., Wallace J.C.;  
RT "A large family of bacterial activator proteins";  
RL Proc. Natl. Acad. Sci. U.S.A. 85(18):6602-6606(1988).  
XX  
RN [6]  
RP 1799-2187  
RX MEDLINE; 85234358.  
RX PUBMED; 3891724.  
RA Haughn G.W., Squires C.H., Defelice M., Largo C.T., Calvo J.M.;  
RT "Unusual organization of the ilvIH promoter of Escherichia coli";  
RL J. Bacteriol. 163(1):186-198(1985).  
XX  
RN [7]  
RP 2101-4431  
RX MEDLINE; 83272971.  
RX PUBMED; 6308579.  
RA Squires C.H., Defelice M., Devereux J., Calvo J.M.;  
RT "Molecular structure of ilvIH and its evolutionary relationship to ilvG in  
RT Escherichia coli";  
RL Nucleic Acids Res. 11(15):5299-5313(1983).  
XX  
RN [8]  
RP 4274-6093  
RX MEDLINE; 90330585.  
XX  
RX PUBMED; 2198273.  
RA Leclerc G., Noel G., Drapeau G.;  
RT "Molecular cloning, nucleotide sequence and expression of shi, a new gene  
RL in the 2-minute region of the genetic map of Escherichia coli";  
XX J. Bacteriol. 172(8):4696-4700(1990).  
RN [9]  
RP 6088-7587  
RX MEDLINE; 90251464.  
RA Gomez M.J., Flouret B., Van Heijenoort J., Ayala J.A.;  
RT "Nucleotide sequence of the regulatory region of pbbB gene of Escherichia  
RL coli";  
XX Nucleic Acids Res. 18:2813-2813(1990).  
RN [10]  
RP 7316-10074  
RX MEDLINE; 83296957.  
RX PUBMED; 6350821.  
RA Nakamura M., Maruyama I.N., Soma M., Kato J.I., Suzuki H., Hirota Y.;  
RT "On the process of cellular division in Escherichia coli: nucleotide  
RL sequence of the gene for penicillin-binding protein 3";  
XX Mol. Gen. Genet. 191(1):1-9(1983).  
RN [11]  
RP 90124047.  
RX MEDLINE; 2692800.  
RA Tao J.S., Ishiguro E.E.;  
RT "Nucleotide sequence of the murE gene of Escherichia coli";  
RL Can. J. Microbiol. 35(11):1051-1054(1989).  
XX  
RN [12]  
RP 90328986.  
RX MEDLINE; 2198024.  
RA Michaud C., Parquet C., Flouret B., Blanot D., Van Heijenoort J.;  
RT "Revised interpretation of the sequence containing the murE gene encoding  
RL the UDP-N-acetylmuramyl-tripeptide synthetase of Escherichia coli";  
XX Biochem. J. 269(1):277-278(1990).  
RN [13]  
RP 11142-12634  
RX MEDLINE; 89345095.  
RA Parquet C., Flouret B., Menguin-Lecreulx D., Van Heijenoort J.;  
RT "Nucleotide sequence of the murF gene encoding the UDP-MurNac- pentapeptide  
RL synthetase of Escherichia coli";  
XX Nucleic Acids Res. 17:5379-5379(1989).  
RN [14]  
RP 12423-15030  
RX MEDLINE; 90192099.  
RA Ikeda M., Wachi M., Ishino F., Matsuhashi M.;  
RT "Nucleotide sequence involving murD and an open reading frame orf-Y spacing  
RL murF and ftsW in Escherichia coli";  
XX Nucleic Acids Res. 18:1058-1058(1990).  
RN [15]  
RP 13392-15020  
RX MEDLINE; 90036736.  
RA Menguin-Lecreulx D., Van Heijenoort J.;  
RT "Nucleotide sequence of the murD gene encoding the UDP-MurNac- L-Ala-D-gly  
RL synthetase of Escherichia coli";  
XX Nucleic Acids Res. 18:183-183(1990).  
RN [16]  
RP 14743-16239  
RX MEDLINE; 2509435.  
RA Ikeda M., Sato T., Wachi M., Jung H.K., Ishino F., Kobayashi Y.,  
RT Matsuhashi M.;  
RL "Structural similarity among Escherichia coli FtsW and RodA proteins and  
RT Bacillus subtilis SpvE protein, which function in cell division, cell  
RT elongation, and spore formation, respectively";  
RL J. Bacteriol. 171(11):6375-6378(1989).  
XX  
RN [17]

RP 16094-17806  
RA Mengin-Lecreux D., Textier L., Van Heijenoort J.;  
RT "Nucleotide sequence of the cell-envelope murC gene of *Escherichia coli*";  
RL Nucleic Acids Res. 18:2810-2810(1990).  
XX [18]  
RN 16094-18886  
RP MEDLINE; 90326550.  
RX Ikeda M., Wachi M., Jung H.K., Ishino F., Matsushashi M.;  
RA "Nucleotide sequence of the cell-envelope murG and murC in the *mra* gene cluster region  
RT of *Escherichia coli*";  
RL Nucleic Acids Res. 18:4014-4014(1990).  
XX [19]  
RN 19464-21952  
RP MEDLINE; 85054557.  
RX PUBMED; 6094474.  
RA Robinson A.C., Kenan D.J., Hatfull G.F., Sullivan N.F., Spiegelberg R.,  
RA Donachie W.D.;  
RT "DNA sequence and transcriptional organization of essential cell division  
RT genes *ftsQ* and *ftsA* of *Escherichia coli*: evidence for overlapping  
RT transcriptional units";  
RL J. Bacteriol. 160(2):546-555(1984).  
XX [20]  
RN 21464-23333  
RP MEDLINE; 86083166.  
RX PUBMED; 3000876.  
RA Yi Q.M., Lutkenhaus J.;  
RT "The nucleotide sequence of the essential cell-division gene *ftsZ* of  
RT *Escherichia coli*";  
RL Gene 36(3):241-247(1985).  
XX [21]  
RN 22964-25011  
RP MEDLINE; 88058745.  
RX PUBMED; 2824434.  
RA Beall B., Lutkenhaus J.;  
RT "Sequence analysis, transcriptional organization, and insertional  
RT mutagenesis of the *envA* gene of *Escherichia coli*";  
RL J. Bacteriol. 169(12):5408-5415(1987).  
XX [22]  
RN 23989-27799  
RP MEDLINE; 88298644.  
RX PUBMED; 2841285.  
RA Schmidt M., Rollo E., Grodberg I., Oliver D.;  
Query Match 1.7%; Score 24; DB 15; Length 28277;  
Best Local Similarity 100.0%; Pred. No. 3.2;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 537 CTGTCGAGCTTCAGCTGGAAC 560  
|||||  
Db 14091 CTGTCGAGCTTCAGCTGGAAC 14114  
ECOL10K 111408 bp DNA linear BCT 29-MAY-2002  
Escherichia coli genomic DNA.(0.0 - 2.4 min).  
D10483 J01597 J01863 J01706 K01298 K01990 M10420 M10611 M12544  
V00259 X04711 X54847 X54945 X55034 X56742  
D10483.2 GI:21321891  
Complete and shotgun sequencing; thrA; thrAl; thrA2; thrB; thrC;  
yaaA; yaaJ; talB; mog; chlG; dnaK; dnaJ; groP; gef; nhaA; ant;  
rpsT; yaaC; ileS; lpsA; yaaB; lytB; yaaF; dgpB; carA; pyrA; carB;  
yaaV; calE; calD; calC; calT; fixA; fixC; yaaT; yabE;  
yabF; kefC; trkC; folA; tmrA; apaH; apaG; pdaA; surA; imp; otaA;  
yabH; yabQ; yabP; yabN; yabM; yabD; leuA; leuB; leuA; leuP; lueO;  
yabJ; yabK; tdpA; yabN; yabM; yabD; leuA; leuB; leuA; leuP; lueO;  
llvI; llvH; brnP; shI; fruR; yabB; yabC; ftsL; pbbP; ftsI; murE;  
murF; mra; mraY; murX; murD; ftsW; murC; ddl; ddb; ftsQ;

SOURCE  
ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

ftsA; divA; ftsZ; sfIB; sulB; lpxC; envA; asmB; yacA; secA; prLD;  
azi; pes; mutT; yacG; htaG; htp; yaaI; nhaR; antO; calF; fixB;  
yaaU; yabO.  
Escherichia coli  
Escherichia coli  
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
Enterobacteriaceae; Escherichia.  
1  
Ohtsubo, H. and Ohtsubo, E.  
Nucleotide sequence of an insertion element, ISI  
Proc. Natl. Acad. Sci. U.S.A. 75 (2), 615-619 (1978)  
78137003  
273224  
2  
Smith, B.R. and Schleif, R.  
Nucleotide sequence of the L-arabinose regulatory region of  
Escherichia coli K12  
J. Biol. Chem. 253 (19), 6931-6933 (1978)  
79005683  
357433  
3  
Greenfield, L., Boone, T. and Wilcox, G.  
DNA sequence of the *araBAD* promoter in *Escherichia coli* B/r  
Proc. Natl. Acad. Sci. U.S.A. 75 (10), 4724-4728 (1978)  
79116194  
368797  
4  
Johnsrud, L.  
DNA sequence of the transposable element ISI  
Mol. Gen. Genet. 169 (2), 213-218 (1979)  
79177885  
375010  
5  
Smith, D.R. and Calvo, J.M.  
Nucleotide sequence of the *E. coli* gene coding for dihydrofolate  
reductase  
Nucleic Acids Res. 8 (10), 2255-2274 (1980)  
81053692  
6159575  
6  
Ogden, S., Haggerty, D., Stoner, C.M., Kolodrubetz, D. and Schleif, R.  
The *Escherichia coli* L-arabinose operon: binding sites of the  
regulatory proteins and a mechanism of positive and negative  
regulation  
Proc. Natl. Acad. Sci. U.S.A. 77 (6), 3346-3350 (1980)  
81013881  
6251457  
7  
Katinka, M., Cossart, P., Sibilli, L., Saint-Girons, I.,  
Chaivignac, M.A., Le Bras, G., Cohen, G.N. and Yaniv, M.  
Nucleotide sequence of the *thrA* gene of *Escherichia coli*  
Proc. Natl. Acad. Sci. U.S.A. 77 (10), 5730-5733 (1980)  
81077247  
7003595  
8  
Miyada, C.G., Horwitz, A.H., Cass, L.G., Tinko, J. and Wilcox, G.  
DNA sequence of the *araC* regulatory gene from *Escherichia coli* B/r  
Nucleic Acids Res. 8 (22), 5267-5274 (1980)  
81124262  
7008027  
9  
Cossart, P., Katinka, M. and Yaniv, M.  
Nucleotide sequence of the *thrB* gene of *E. coli*, and its two  
adjacent regions; the *thrAB* and *thrBC* junctions  
Nucleic Acids Res. 9 (2), 339-347 (1981)  
81150470  
6259626  
10  
Lee, N.L., Gielow, W.O. and Wallace, R.G.  
Mechanism of *araC* autoregulation and the domains of two overlapping  
promoters, *PC* and *PBAD*, in the L-arabinose regulatory region of  
*Escherichia coli*  
Proc. Natl. Acad. Sci. U.S.A. 78 (2), 752-756 (1981)



MEDLINE  
 PUBMED  
 REFERENCE  
 11  
 AUTHORS  
 TITLE  
 Nucleotide sequence of the gene for ribosomal protein S20 and its  
 flanking regions  
 J. Biol. Chem. 256 (15), 8177-8182 (1981)  
 MEDLINE  
 PUBMED  
 REFERENCE  
 12  
 AUTHORS  
 TITLE  
 Is the amino acid but not the nucleotide sequence of the  
 Escherichia coli *araC* gene conserved?  
 J. Mol. Biol. 154 (4), 649-652 (1982)  
 MEDLINE  
 PUBMED  
 REFERENCE  
 13  
 AUTHORS  
 TITLE  
 Sequence of the *malK* gene in *E. coli* K12  
 Nucleic Acids Res. 10 (22), 7449-7458 (1982)  
 MEDLINE  
 PUBMED  
 REFERENCE  
 14  
 AUTHORS  
 TITLE  
 Sequence of the *thrC* and of the transcription termination  
 region of the threonine operon in *Escherichia coli* K12  
 Nucleic Acids Res. 11 (21), 7331-7345 (1983)  
 MEDLINE  
 PUBMED  
 REFERENCE  
 15  
 AUTHORS  
 TITLE  
 Major heat shock gene of *Drosophila* and the *Escherichia coli*  
 heat-inducible *dnkA* gene are homologous  
 Proc. Natl. Acad. Sci. U.S.A. 81 (3), 848-852 (1984)  
 MEDLINE  
 PUBMED  
 REFERENCE  
 16  
 AUTHORS  
 TITLE  
 Nucleotide sequence of the *Escherichia coli* prolipoprotein signal  
 peptidase (*lsp*) gene  
 Proc. Natl. Acad. Sci. U.S.A. 81 (12), 3708-3712 (1984)  
 MEDLINE  
 PUBMED  
 REFERENCE  
 17  
 AUTHORS  
 TITLE  
 Multiple regulatory signals in the control region of the  
*Escherichia coli* *carAB* operon  
 Proc. Natl. Acad. Sci. U.S.A. 81 (13), 4139-4143 (1984)  
 MEDLINE  
 PUBMED  
 REFERENCE  
 18  
 AUTHORS  
 TITLE  
 Nucleotide sequence and expression of the *Escherichia coli* *dapB*  
 gene  
 J. Biol. Chem. 259 (23), 14829-14834 (1984)  
 MEDLINE  
 PUBMED  
 REFERENCE  
 19  
 AUTHORS  
 TITLE  
 Characterization of the *ilvS-lsp* operon in *Escherichia coli*.  
 Identification of an open reading frame upstream of the *ilvS* gene  
 and potential promoter(s) for the *ilvS-lsp* operon  
 J. Biol. Chem. 260 (9), 5616-5620 (1985)  
 MEDLINE  
 PUBMED  
 REFERENCE  
 20  
 AUTHORS  
 TITLE  
 Characterization of the 3' end of the leucine operon of *Salmonella*  
 typhimurium  
 Mol. Gen. Genet. 199 (3), 486-494 (1985)  
 MEDLINE  
 PUBMED  
 REFERENCE  
 21  
 AUTHORS  
 TITLE  
 Consensus sequence for *Escherichia coli* heat shock gene promoters  
 Proc. Natl. Acad. Sci. U.S.A. 82 (9), 2679-2683 (1985)  
 MEDLINE  
 PUBMED  
 REFERENCE  
 22  
 AUTHORS  
 TITLE  
 Structural analysis of a new GC-specific insertion element IS186  
 FEBS Lett. 192 (1), 47-52 (1985)  
 MEDLINE  
 PUBMED  
 REFERENCE  
 23  
 AUTHORS  
 TITLE  
 Cloning and characterization of a cDNA encoding the rat brain  
 glucose-transporter protein  
 Proc. Natl. Acad. Sci. U.S.A. 83 (16), 5784-5788 (1986)  
 MEDLINE  
 PUBMED  
 REFERENCE  
 24  
 AUTHORS  
 TITLE  
 DNA and amino-acid sequences of 3-isopropylmalate dehydrogenase of  
*Bacillus coagulans*. Comparison with the enzymes of *Saccharomyces*  
*cerevisiae* and *Thermus thermophilus*  
 Biochim. Biophys. Acta 867, 36-44 (1986)  
 MEDLINE  
 PUBMED  
 REFERENCE  
 25  
 AUTHORS  
 TITLE  
 Nucleotide sequence of the *Escherichia coli* *dnaJ* gene and  
 purification of the gene product  
 J. Biol. Chem. 261 (4), 1778-1781 (1986)  
 MEDLINE  
 PUBMED  
 REFERENCE  
 26  
 AUTHORS  
 TITLE  
 Structure of the DNA distal to the gene for ribosomal protein S20  
 in *Escherichia coli* K12: presence of a strong terminator and an IS1  
 element  
 Nucleic Acids Res. 14 (17), 6965-6981 (1986)  
 MEDLINE  
 PUBMED  
 REFERENCE  
 27  
 AUTHORS  
 TITLE  
 Query Match. 1.7%; Score 24; DB 1; Length 111408;  
 Best Local Similarity 100.0%; Pred. NO. 2.6;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 537 CTCTCGAGCTCCAGCTGGAACCC 560  
 Db 97222 CTCTCGAGCTCCAGCTGGAACCC 97245  
 RESULT 13  
 AC073937  
 LOCUS  
 DEFINITION  
 Mus musculus clone RP23-43M3 strain C57BL6/J, WORKING DRAFT  
 SEQUENCE, 44 unordered pieces.  
 AC073937  
 AC073937.14 GI:14269588  
 HTG: HTGS PHASE1: HTGS DRAFT.  
 SOURCE  
 Mus musculus (house mouse)  
 ORGANISM  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 248308)  
 Montgomerly, K.T., Grills, G., Han, J., Lee, E., Long, J., Pomerantz, R.,  
 Toshikhes, I.P., Shim, C., Decker, J., Thomas, E., Perera, A.,  
 Gordon, M., Goltz, J.S. and Kuchlerlapati, R.  
 High Throughput Mouse Sequencing  
 Unpublished  
 2 (bases 1 to 248308)  
 Montgomerly, K.T., Grills, G., Han, J., Lee, E., Long, J., Pomerantz, R.,  
 Toshikhes, I.P., Shim, C., Decker, J., Thomas, E., Perera, A.,  
 Gordon, M., Goltz, J.S., Liedtke, W., Friedman, J.M. and  
 Kuchlerlapati, R.

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 21  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 22  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 23  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 24  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 25  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 26  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 27  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 28  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 29  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 30  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 31  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 32  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 33  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 34  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 35  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 36  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 37  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 38  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 39  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 40  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 41  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 42  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 43  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 44  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 45  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 46  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 47  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 48  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 49  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 50  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 51  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 52  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 53  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 54  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 55  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 56  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 57  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 58  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 59  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 60  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 61  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 62  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 63  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 64  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 65  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 66  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 67  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 68  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 69  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 70  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 71  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 72  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 73  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 74  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 75  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 76  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 77  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 78  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 79  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 80  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 81  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 82  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 83  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 84  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 85  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 86  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 87  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 88  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 89  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 90  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 91  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 92  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 93  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 94  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 95  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 96  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 97  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 98  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 99  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 100  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE

TITLE Direct Submission  
 JOURNAL Submitted (07-JUL-2000) Department of Molecular Genetics, Albert Einstein College of Medicine Genome Center, 1300 Morris Park Ave., Bronx, NY 10461, USA  
 COMMENT On May 31, 2001 this sequence version replaced gi:14190542.  
 -----Genome Center  
 Center: Harvard Partners Genome Center  
 Center Code: HPGC  
 Web site: <http://www.hpgc.org/Sequence/mouse.html>  
 Contact: hpgc@medel.mgh.harvard.edu  
 -----Summary Statistics  
 Center project name: AEX  
 Sequencing vector: pUC18; L08752  
 Chemistry: Dye-terminator Big Dye; 100%  
 \*Consensus quality: 231072 at least Q20  
 \*Consensus quality: 226894 at least Q30  
 \*Consensus quality: 220122 at least Q40  
 Estimated insert size: agarose-FP - N/A  
 \*Estimated insert size: 247448 - sum-of-contigs  
 Quality coverage: agarose-FP - N/A  
 Quality coverage: 7.4 x in Q20 bases; sum-of-contigs estimation  
 -----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 44 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 -----  
 1 45431: contig of 45431 bp in length  
 \* 45432 45451: gap of unknown length  
 \* 45452 90348: contig of 44897 bp in length  
 \* 90349 90368: gap of unknown length  
 \* 90369 119208: contig of 28840 bp in length  
 \* 119209 119228: gap of unknown length  
 \* 119229 134414: contig of 15186 bp in length  
 \* 134415 134434: gap of unknown length  
 \* 134435 149784: contig of 15350 bp in length  
 \* 149785 149804: gap of unknown length  
 \* 149805 158841: contig of 9037 bp in length  
 \* 158842 158861: gap of unknown length  
 \* 158862 166440: contig of 7579 bp in length  
 \* 166441 166460: gap of unknown length  
 \* 166461 176603: contig of 10143 bp in length  
 \* 176604 176623: gap of unknown length  
 \* 176624 185346: contig of 8723 bp in length  
 \* 185347 185366: gap of unknown length  
 \* 185367 192080: contig of 6714 bp in length  
 \* 192081 192100: gap of unknown length  
 \* 192101 199361: contig of 7261 bp in length  
 \* 199362 199381: gap of unknown length  
 \* 199382 204102: contig of 4721 bp in length  
 \* 204103 204122: gap of unknown length  
 \* 204123 207075: contig of 2953 bp in length  
 \* 207076 207095: gap of unknown length  
 \* 207096 209883: contig of 2788 bp in length  
 \* 209884 209903: gap of unknown length  
 \* 209904 211730: contig of 1827 bp in length  
 \* 211731 211750: gap of unknown length  
 \* 211751 213370: contig of 1620 bp in length  
 \* 213371 213390: gap of unknown length  
 \* 213391 215364: contig of 1974 bp in length  
 \* 215365 215384: gap of unknown length  
 \* 215385 216640: contig of 1256 bp in length  
 \* 216641 216660: gap of unknown length  
 \* 216661 217471: contig of 811 bp in length  
 \* 217472 217491: gap of unknown length  
 \* 217492 220033: contig of 2542 bp in length  
 \* 220034 220053: gap of unknown length  
 \* 220054 221452: contig of 1399 bp in length  
 \* 221453 221472: gap of unknown length  
 \* 221473 223540: contig of 2068 bp in length

misc\_feature /note="assembly\_name:Contig195"  
192101. .199361  
misc\_feature /note="assembly\_name:Contig194"  
199382. .204102  
misc\_feature /note="assembly\_name:Contig193"  
204123. .207075  
misc\_feature /note="assembly\_name:Contig192"  
207096. .209883  
misc\_feature /note="assembly\_name:Contig191"  
209904. .211730  
/note="assembly\_name:Contig190  
clone\_end:SP6  
vector\_side:left"  
misc\_feature 211751. .213370  
/note="assembly\_name:Contig189"  
213391. .215364  
misc\_feature /note="assembly\_name:Contig188"  
215385. .216640  
misc\_feature /note="assembly\_name:Contig187"  
216661. .217471  
/note="assembly\_name:Contig186"  
217492. .220033  
misc\_feature /note="assembly\_name:Contig185"  
220054. .221452  
/note="assembly\_name:Contig184"  
221473. .223540  
misc\_feature /note="assembly\_name:Contig183"  
223561. .224311  
/note="assembly\_name:Contig182"  
224332. .224943  
misc\_feature /note="assembly\_name:Contig181"  
224964. .226352  
/note="assembly\_name:Contig180"

Query Match 1.7%; Score 24; DB 2; Length 248308;

Best Local Similarity 100.0%; Pred. No. 2.3;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 537 CTGTCGAGCTTCCAGCTGGAACC 560  
|||||  
Db 239530 CTGTCGAGCTTCCAGCTGGAACC 239553

RESULT 14  
AC020876/c  
LOCUS AC020876 259498 bp DNA linear HTG 29-MAR-2000  
DEFINITION Mus musculus clone RP23-336D4, WORKING DRAFT SEQUENCE, 75 unordered  
pieces.  
ACCESSION AC020876  
VERSION AC020876.2 GI:7340302  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 259498)  
AUTHORS DOE Joint Genome Institute.  
TITLE Sequencing of Mouse  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 259498)  
AUTHORS DOE Joint Genome Institute.  
TITLE Direct Submission  
JOURNAL Submitted (10-JAN-2000) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
COMMENT On Mar 29, 2000 this sequence version replaced gi:6686432.  
-----Genome Center  
Center: Joint Genome Institute  
Center Code: JGI  
Web site: <http://www.jgi.doe.gov>  
-----Summary Statistics  
Consensus quality: 178059 bases at least Q40  
Consensus quality: 214074 bases at least Q30  
Consensus quality: 225088 bases at least Q20

Estimated insert size: 259498; sum-of-contigs estimation  
Estimated insert size: 206000; pulse field gel estimation  
Quality coverage: 4.28x in Q20 bases; pulse-field gel estimation  
Quality coverage: 3.40x in Q20 bases; sum-of-contigs estimation  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 75 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 1047: contig of 1047 bp in length  
\* \* 1048 gap of unknown length  
\* \* 2408: contig of 1361 bp in length  
\* \* gap of unknown length  
\* \* 2409 contig of 1013 bp in length  
\* \* gap of unknown length  
\* \* 3422 contig of 1322 bp in length  
\* \* gap of unknown length  
\* \* 4744 contig of 1040 bp in length  
\* \* gap of unknown length  
\* \* 5784 contig of 1023 bp in length  
\* \* gap of unknown length  
\* \* 6807 contig of 1743 bp in length  
\* \* gap of unknown length  
\* \* 8550 contig of 1488 bp in length  
\* \* gap of unknown length  
\* \* 10038 contig of 1129 bp in length  
\* \* gap of unknown length  
\* \* 11167 contig of 1171 bp in length  
\* \* gap of unknown length  
\* \* 12338 contig of 1065 bp in length  
\* \* gap of unknown length  
\* \* 13403 contig of 1151 bp in length  
\* \* gap of unknown length  
\* \* 14554 contig of 1095 bp in length  
\* \* gap of unknown length  
\* \* 15649 contig of 1675 bp in length  
\* \* gap of unknown length  
\* \* 17324 contig of 1123 bp in length  
\* \* gap of unknown length  
\* \* 18447 contig of 1835 bp in length  
\* \* gap of unknown length  
\* \* 20282 contig of 1370 bp in length  
\* \* gap of unknown length  
\* \* 21652 contig of 1128 bp in length  
\* \* gap of unknown length  
\* \* 22780 contig of 1049 bp in length  
\* \* gap of unknown length  
\* \* 23829 contig of 1141 bp in length  
\* \* gap of unknown length  
\* \* 24970 contig of 1631 bp in length  
\* \* gap of unknown length  
\* \* 26601 contig of 1165 bp in length  
\* \* gap of unknown length  
\* \* 27766 contig of 1437 bp in length  
\* \* gap of unknown length  
\* \* 29203 contig of 1436 bp in length  
\* \* gap of unknown length  
\* \* 30639 contig of 1712 bp in length  
\* \* gap of unknown length  
\* \* 32351 contig of 1464 bp in length  
\* \* gap of unknown length  
\* \* 33815 contig of 1179 bp in length  
\* \* gap of unknown length  
\* \* 34994 contig of 1379 bp in length  
\* \* gap of unknown length  
\* \* 36373 contig of 1339 bp in length  
\* \* gap of unknown length  
\* \* 37712 contig of 1110 bp in length  
\* \* gap of unknown length

\* 38822 40400: contig of 1579 bp in length  
\* gap of unknown length  
\* 40401 41429: contig of 1029 bp in length  
\* gap of unknown length  
\* 41430 42543: contig of 1114 bp in length  
\* gap of unknown length  
\* 42544 43957: contig of 1414 bp in length  
\* gap of unknown length  
\* 43958 45245: contig of 1288 bp in length  
\* gap of unknown length  
\* 45246 47013: contig of 1768 bp in length  
\* gap of unknown length  
\* 47014 48155: contig of 1142 bp in length  
\* gap of unknown length  
\* 48156 49158: contig of 1003 bp in length  
\* gap of unknown length  
\* 49159 50170: contig of 1012 bp in length  
\* gap of unknown length  
\* 50171 51392: contig of 1222 bp in length  
\* gap of unknown length  
\* 51393 53265: contig of 1873 bp in length  
\* gap of unknown length  
\* 53266 55279: contig of 2014 bp in length  
\* gap of unknown length  
\* 55280 56908: contig of 1629 bp in length  
\* gap of unknown length  
\* 56909 58527: contig of 1619 bp in length  
\* gap of unknown length  
\* 58528 60093: contig of 1566 bp in length  
\* gap of unknown length  
\* 60094 61444: contig of 1351 bp in length  
\* gap of unknown length  
\* 61445 62672: contig of 1228 bp in length  
\* gap of unknown length  
\* 62673 65545: contig of 2873 bp in length  
\* gap of unknown length  
\* 65546 67540: contig of 1995 bp in length  
\* gap of unknown length  
\* 67541 68584: contig of 1044 bp in length  
\* gap of unknown length  
\* 68585 71572: contig of 2988 bp in length  
\* gap of unknown length  
\* 71573 73312: contig of 1740 bp in length  
\* gap of unknown length  
\* 73313 75668: contig of 2356 bp in length  
\* gap of unknown length  
\* 75669 78367: contig of 2699 bp in length  
\* gap of unknown length  
\* 78368 80669: contig of 2302 bp in length  
\* gap of unknown length  
\* 80670 83363: contig of 2694 bp in length  
\* gap of unknown length  
\* 83364 87042: contig of 3679 bp in length  
\* gap of unknown length  
\* 87043 89727: contig of 2685 bp in length  
\* gap of unknown length  
\* 89728 93097: contig of 3370 bp in length  
\* gap of unknown length  
\* 93098 96142: contig of 3045 bp in length  
\* gap of unknown length  
\* 96143 99546: contig of 3404 bp in length  
\* gap of unknown length  
\* 99547 105084: contig of 5538 bp in length  
\* gap of unknown length  
\* 105085 109494: contig of 4410 bp in length  
\* gap of unknown length  
\* 109495 113825: contig of 4331 bp in length  
\* gap of unknown length  
\* 113826 119314: contig of 5489 bp in length  
\* gap of unknown length  
\* 119315 129224: contig of 9910 bp in length  
\* gap of unknown length  
\* 129225 136674: contig of 7450 bp in length

\* 136675 144863: gap of unknown length  
\* contig of 8189 bp in length  
\* 144864 155259: gap of unknown length  
\* contig of 10396 bp in length  
\* 155260 164645: gap of unknown length  
\* contig of 9386 bp in length  
\* 164646 174742: gap of unknown length  
\* contig of 10097 bp in length  
\* 174743 184830: gap of unknown length  
\* contig of 10088 bp in length  
\* 184831 201876: gap of unknown length  
\* contig of 17046 bp in length  
\* 201877 221979: gap of unknown length  
\* contig of 20103 bp in length  
\* 221980 259498: gap of unknown length  
\* contig of 37519 bp in length.

FEATURES  
source  
Location/Qualifiers  
1. 259498  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/clone="Rp23-336D4"

BASE COUNT 72952 a 58720 c 57559 g 70128 t 139 others  
ORIGIN

Query Match 1.7%; Score 24; DB 2; Length 259498;  
Best Local Similarity 100.0%; Pred. No. 2.3;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 537 CTGTCGAGCTTCAGCTGGAACCC 560  
|||||  
Db 50010 CTGTCGAGCTTCAGCTGGAACCC 49987

RESULT 15  
LOCUS AP002550 281530 bp DNA linear BCT 07-MAR-2001  
DEFINITION Escherichia coli O157:H7 DNA, complete genome, section 1/20.  
ACCESSION AP002550 BA000007  
VERSION AP002550.1 GI:13359456  
KEYWORDS  
SOURCE Escherichia coli O157:H7  
ORGANISM Escherichia coli O157:H7  
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
Enterobacteriaceae; Escherichia.  
REFERENCE 1 (sites)  
AUTHORS Makino,K., Yokoyama,K., Kubota,Y., Yutsudo,C.H., Kimura,S.,  
Kurokawa,K., Ishii,K., Hattori,M., Tatsuno,I., Abe,H., Iida,T.,  
Yamamoto,K., Ohnishi,M., Hayashi,T., Yasunaga,T., Honda,T.,  
Sasakawa,C. and Shinagawa,H.  
Complete nucleotide sequence of the prophage VT2-Sakai carrying the  
verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7  
derived from the Sakai outbreak  
Genes Genet. Syst. 74 (5), 227-239 (1999)

JOURNAL 20198780  
MEDLINE 10734605  
PUBMED  
REFERENCE 2 (sites)  
AUTHORS Ohnishi,M., Murata,T., Nakayama,K., Kuhara,S., Hattori,M.,  
Kurokawa,K., Yasunaga,T., Yokoyama,K., Makino,K., Shinagawa,H. and  
Hayashi,T.  
Comparative analysis of the whole set of rRNA operons between an  
enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an  
Escherichia coli K-12 strain MG1655  
Syst. Appl. Microbiol. 23 (3), 315-324 (2000)

JOURNAL 20557356  
MEDLINE 11108008  
PUBMED  
REFERENCE 3 (sites)  
AUTHORS Yokoyama,K., Makino,K., Kubota,Y., Watanabe,M., Kimura,S.,  
Yutsudo,C.H., Kurokawa,K., Ishii,K., Hattori,M., Abe,H., Iida,T.,  
Yamamoto,K., Hayashi,T., Yasunaga,T., Honda,T., Sasakawa,C. and  
Shinagawa,H.  
Complete nucleotide sequence of the prophage VT1-Sakai carrying the  
Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli

JOURNAL	0157:H7 strain derived from the Sakai outbreak	gene
MEDLINE	Gene 258 (1-2), 127-139 (2000)	CDS
PUBMED	20564182	
REFERENCE	11111050	
AUTHORS	4 (sites) Hayashi,T., Makino,K., Ohnishi,M., Kurokawa,K., Ishii,K., Yokoyama,K., Han,C.-G., Ohtsubo,E., Nakayama,K., Murata,T., Tanaka,M., Tobe,T., Iida,T., Takami,H., Honda,T., Sasakawa,C., Ogasawara,N., Yasunaga,T., Kuhara,S., Shiba,T., Hattori,M. and Shinagawa,H. Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic comparison with a laboratory strain K-12 DNA Res. 8 (1), 11-22 (2001) 21156231 11258796 5 (bases 1 to 281530) Ohnishi,M., Kurokawa,K., Makino,K., Yasunaga,T., Shinagawa,H. and Hayashi,T. Direct Submission Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871, Japan (E-mail:ken@gen-info.osaka-u.ac.jp, URL:http://www.gen-info.osaka-u.ac.jp/, Tel:81-6-6879-8365, Fax:81-6-6879-2047) genome project. Location/Qualifiers 1..281530 /organism="Escherichia coli O157:H7" /mol_type="genomic DNA" /strain="O157:H7" /sub_strain="RIMD 0509952" /db_xref="taxon:83334" 190..273 /gene="ECs0001" 190..273 /gene="ECs0001" /note="similar to THRL_ECOLI gi 1786182 percent similarity 100 in 21 aa, but has 6 additional residues (Conserved in E.coli K-12)" /codon_start=1 /evidence=not_experimental /transl_table=11 /product="thr operon leader peptide" /protein_id="BAB33424.1" /db_xref="GI:13359457" /translation="MKRISTTITTTTITTTITTTITTTGNGAG" 354..2816 /gene="ECs0002" 354..2816 /gene="ECs0002" /note="similar to THRA_ECOLI gi 1786183 percent identity 99 in 820 aa (Conserved in E.coli K-12)" /codon_start=1 /evidence=not_experimental /transl_table=11 /product="aspartokinase I-homoserine dehydrogenase I" /protein_id="BAB33425.1" /db_xref="GI:13359458" /translation="NRVLKFGGTSVANAERFLRVADITLESNARQGVATVLSAPAKIT NHLVAMIEKTSQQDALPNISDAERIFAEITGLAAQPGFPLAQKITFVDQEFQAIK HVLHGILSGQCPDSINAAICRGKMSIAIMAGVLEARGHNVTVIDPVEKLAVGHY LESTVDIAESTRIRTAASRIPADHVMLAGFTAGNEKGLSVYVGLRNGSDYSAVLAACL RADCEIWDVDGYTCDPDPDARILKMSYOFAMELSYFGAGVLHPRTITPTAOF OIPCLIKRTGNPQAGTLIGASRDEDEL PVKGI SNLNNAMESVSGPGMKMGVGAAR VFAMASRVLVLTQSSEYSISFCVPQSDCVRAAMQEEYFLEKGLGEPVLAV TERLAIISVGDGMTLRGISAKFAALARANINIAAQQSSERSISVNNDDATT GVRVTHQMLFNTQDIEYFVIGVGVGALLQKROOSWLKNKHIDRLRVCGVANSKA LITNVHGLNENQWBEAQAKPEPNLGRILRLVKEYHLLNPEYIVDCTSQAQVADYAD FLGEGHVVTNKNKNTSSMDYHLLRHAAEKSRKFLYDYNVAGLPLVINLQNLN AGDELMKESGILSGLSYIFGKLDEGMSFEATFLAREMGTYTEPDRDLGSMOVARK LLILARETGLELELADIEIEPVLPAEFNAGDVAFMANLSQDLDFARVAKMADEG KVLRYVGNIDEGVGRVKIAEVDGNDPLFKVKNGENALAFYSHYQYPLPLVLRGYGAG NDVTAAGVFADLLRTLSWKLVG"	
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene
JOURNAL		CDS
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		gene

/note="similar to YAAJ\_ECOLI gi11786188 percent identity  
99 in 476 aa (Conserved in E.coli K-12)"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11  
/product="putative inner membrane transport protein"  
/protein\_id="BAB33430.1"  
/db\_xref="GI:13359463"  
/translation="MPDFFSFINSVLWGSVMYLLFGAGCWETFRGTGFVQFVIRQFG  
KSLKNSITHPOPGGLTSFQSLCTSLAARVSGSNLAGVALAITAGPGGAVFWMWVAEFIG  
MATSEKCSLAQLYKERDVGQPRGGPAWYMARGLMRWGVLFVFLLIAYGIIFSG  
VQANAVARALUSFDFPPLVTGIILAVFALLIIRGLHGVARLMQGFVPLMAIIWVLT  
SLVICVINIGQLPHVINSIFESAFGQERAGGAAGYTLQAITNGFQRFMSNEAGMG  
STPNAASAAASWPHPAAGQIVQMIGTIFDTLVICTASAMLIILAGNGTTPMPLEGTQ  
LIQKAMRVLMGSGAEFTLVWILFAFSIVANVIVAENNLFFLRLLNNPKAIWCLRIC  
TEATVIGGTLISLPLMWQADIIMACMAITNLTAILLLSPVHTIASDYLQRKLGVR  
PVFDPLRYPDIGRQLSRDWDVDSQ"

Query Match 1.7%; Score 24; DB 1; Length 281530;  
Best Local Similarity 100.0%; Pred. No. 2.3;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 537 CTGTCGAGCTCCAGCTGGAACC 560  
Db 102166 CTGTCGAGCTCCAGCTGGAACC 102189  
|||||

Search completed: August 14, 2003, 06:15:42  
Job time : 5311 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2003, 01:08:54 ; Search time 5305 Seconds  
(without alignments)  
11181.708 Million cell updates/sec

Title: US-09-701-229-1  
Perfect score: 1450  
Sequence: 1 cgtgctgctgctgctgcca.....tggtgacgcggcagcgcatc 1450

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0  
Searched: 2888711 seqs, 2045481386 residues  
Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb.ba.\*
- 2: gb.htg.\*
- 3: gb.in.\*
- 4: gb.on.\*
- 5: gb.ov.\*
- 6: gb.pat.\*
- 7: gb.ph.\*
- 8: gb.pl.\*
- 9: gb.pr.\*
- 10: gb.ro.\*
- 11: gb.sts.\*
- 12: gb.sy.\*
- 13: gb.vi.\*
- 14: gb.vi.\*
- 15: em.ba.\*
- 16: em.fun.\*
- 17: em.in.\*
- 18: em.in.\*
- 19: em.mu.\*
- 20: em.om.\*
- 21: em.or.\*
- 22: em.ov.\*
- 23: em.pat.\*
- 24: em.ph.\*
- 25: em.pl.\*
- 26: em.ro.\*
- 27: em.sts.\*
- 28: em.un.\*
- 29: em.vi.\*
- 30: em.htg\_hum.\*
- 31: em.htg\_inv.\*
- 32: em.htg\_other.\*
- 33: em.htg\_mus.\*
- 34: em.htg\_pln.\*
- 35: em.htg\_rod.\*
- 36: em.htg\_mam.\*
- 37: em.htg\_vrt.\*
- 38: em\_sy.\*
- 39: em\_htgo\_hum.\*
- 40: em\_htgo\_mus.\*
- 41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
C 1	1449	99.9	24000	1	AE004856	AE004856 Pseudomon
	1396.2	96.3	5280	1	AY008276	AY008276 Pseudomon
C 3	895.4	61.8	301995	1	AE016779	AE016779 Pseudomon
C 4	827.8	57.1	311600	1	AE016871	AE016871 Pseudomon
C 5	335.8	23.2	251050	1	AL627265	AL627265 Salmonell
C 6	335.8	23.2	300169	1	AE016834	AE016834 Salmonell
C 7	332.6	22.9	22348	1	AE008699	AE008699 Salmonell
C 8	291.2	20.1	11441	1	AE015046	AE015046 Shigella
C 9	291.2	20.1	290029	1	AE016978	AE016978 Shigella
C 10	288	19.9	12791	1	AE005185	AE005185 Escherich
C 11	288	19.9	281530	1	AP002550	AP002550 Escherich
C 12	286.4	19.8	300409	1	AE016755	AE016755 Escherich
C 13	282	19.4	2608	1	ECMUROY	X51584 Escherichia
C 14	282	19.4	21757	1	AE000118	AE000118 Escherich
C 15	282	19.4	28277	6	AX191720	AX191720 Sequence
C 16	282	19.4	28277	15	EC2MIN	X55034 E. coli 2 m
C 17	282	19.4	111408	1	ECO110K	D10483 Escherichia
C 18	278.6	19.2	1317	6	AX189058	AX189058 Sequence
C 19	272.4	18.8	1629	1	ECMURD	X17609 Escherichia
C 20	267.6	18.5	188050	1	AL646072	AL646072 Ralstonia
C 21	263.8	18.2	248308	2	AC073937	AC073937 Mus muscu
C 22	259.8	17.9	11726	1	AE013965	AE013965 Yersinia
C 23	253.6	17.5	301050	1	AJ414143	AJ414143 Yersinia
C 24	253.6	17.5	301727	1	AE016960	AE016960 Coxsella
C 25	212.6	14.7	13071	1	AE015855	AE015855 Shevanel
C 26	212.6	14.6	259498	2	AC020876	AC020876 Mus muscu
C 27	194	13.4	316050	1	BX321859	BX321859 Nitrosomo
C 28	191.2	13.2	195767	1	NMA722491	AL627358 Neisseria
C 29	185	12.8	22201	1	AE004310	AE004310 Vibrio ch
C 30	184.6	12.7	298900	1	AP005074	AP005074 Streptomy
C 31	180.6	12.5	299925	1	AP005045	AP005045 Streptomy
C 32	180	12.4	10629	1	AE002398	AE002398 Neisseria
C 33	180	12.4	349980	6	AX044029	AX044029 Sequence
C 34	174.2	12.0	301442	1	AE016798	AE016798 Vibrio vu
C 35	173.8	12.0	37586	1	MTCY270	295388 Mycobacteri
C 36	173.8	12.0	37586	6	AX191745	AX191745 Sequence
C 37	172.2	11.9	20400	1	AE007068	AE007068 Mycobacte
C 38	172.2	11.9	306050	1	BX248341	BX248341 Mycobacte
C 39	160.2	11.0	321250	1	SC093911	AL939111 Streptomy
C 40	155.2	10.7	12401	1	AE005924	AE005924 Caulobact
C 41	152	10.5	12232	1	AE006048	AE006048 Pasteurel
C 42	149.8	10.3	11498	1	U32793	U32793 Haemophilus
C 43	149.8	10.3	11498	6	AX191762	AX191762 Sequence
C 44	149.8	10.3	110000	6	AR274513	Continuation (12 o
C 45	149.8	10.3	110000	6	AR274513_12	Continuation (13 o

ALIGNMENTS

RESULT 1  
AE004856/c  
LOCUS  
DEFINITION Pseudomonas aeruginosa PAO1, section 417 of 529 of the complete genome.  
ACCESSION AE004856 AE004091  
VERSION AE004856.1 GI:9950633  
KEYWORDS Pseudomonas aeruginosa PAO1  
SOURCE Pseudomonas aeruginosa PAO1  
ORGANISM Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
Pseudomonadaceae; Pseudomonas.  
REFERENCE 1 (bases 1 to 24000)  
AUTHORS Stover,C.K., Pham,X.-Q., T., Erwin,A.L., Mizoguchi,S.D., Warrenner,P., Hickey,M.J., Brinkman,F.S.L., Hufnagle,W.O., Kowalik,D.J.,

Lagrou, M., Garber, R.L., Goltry, L., Tolentino, E., Westbrock-Wadman, S., Yuan, Y., Brody, L.L., Coulter, S.N., Folger, K.R., Kas, A., Larbig, K., Lim, R.M., Smith, K.A., Spencer, D.H., Wong, G.K.-S., Wu, Z., and Paulsen, I.T.  
Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen  
Nature 406 (6799), 959-964 (2000)  
2043737  
10984043  
2 (bases 1 to 24000)  
Stover, C.K., Pham, X.-Q.T., Erwin, A.L., Mizoguchi, S.D., Warren, P., Hickey, M.J., Brinkman, F.S.L., Huftnagle, W.O., Kowalik, D.J., Lagrou, M., Garber, R.L., Goltry, L., Tolentino, E., Westbrock-Wadman, S., Yuan, Y., Brody, L.L., Coulter, S.N., Folger, K.R., Kas, A., Larbig, K., Lim, R.M., Smith, K.A., Spencer, D.H., Wong, G.K.-S., Wu, Z., Paulsen, I.T., Reizer, J., Saier, M.H., Hancock, R.E.W., Lory, S. and Olson, M.V.  
Direct Submission  
Submitted (16-MAY-2000) Department of Medicine and Genetics, University of Washington Genome Center, University Of Washington, Box 352145, Seattle, WA 98195, USA  
3 (bases 1 to 24000)

*Pseudomonas aeruginosa* Community Annotation Project (*PseudoCAP*)  
Direct Submission  
Submitted (04-FEB-2003) Department of Molecular Biology and Biochemistry, Simon Fraser University, 8888 University Dr., Burnaby, British Columbia V5A 1S6, Canada

-----  
This represents the February 3, 2003 version of the continually updated, reviewed, *Pseudomonas aeruginosa* PA01 genome annotation, from *PseudoCAP* (see <http://www.pseudomonas.com> for latest updates and links to alternate annotations). *PseudoCAP* is coordinated by Fiona S.L. Brinkman (Simon Fraser University, Canada) and Robert E.W. Hancock (University of British Columbia, Canada). We welcome submission through [www.pseudomonas.com](http://www.pseudomonas.com) of any proposed changes.

'Protein name confidence' is used to rate our confidence of the accuracy of the protein name.  
Class 1: Function experimentally demonstrated in *P. aeruginosa*.  
Class 2: Function of highly similar gene experimentally demonstrated in another organism (and gene context consistent in terms of pathways its involved in, if known).  
Class 3: Function proposed based on presence of conserved amino acid motif, structural feature or limited sequence similarity to an experimentally studied gene.  
Class 4: Homologs of previously reported genes of unknown function, or no similarity to any previously reported sequences.

-----  
Location/Qualifiers  
1. .24000  
/organism="Pseudomonas aeruginosa PA01"  
/mol\_type="genomic DNA"  
/strain="PA01"  
/db\_xref="taxon:208964"  
complement(91. .1308)  
/gene="argJ"  
/locus\_tag="PA4402"  
complement(91. .1308)  
/gene="argJ"  
/locus\_tag="PA4402"  
/note="Protein name confidence: Class 2 (high similarity to functionally studied protein); Subcellular localization: Cytoplasmic; Subcellular localization confidence: Class 2; Alt protein name: ornithine transacetylase"  
/codon\_start=1  
/transl\_table=11  
/product="glutamate N-acetyltransferase"  
/protein\_id="AAG07790.1"  
/db\_xref="GI:9950634"  
/translation="MAVGLGPLSTLHPVPGFELGTASGIKRPGKDVVVMRCAGSS VAGVFTLNFCAAPVTLAKQRFELGEVRYLLTWTGNANAGTGEAGLAAAQTCAKLAEL

## TITLE

JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS

## TITLE

JOURNAL  
AUTHORS  
CONSTRM  
TITLE  
JOURNAL

## COMMENT

FEATURES  
source

## gene

## CDS

AGVAETSVLPYSTGVIGEPVPAKIEAALPAALADLAEDRWAEAAAGIMTTDTLPKGA  
SRQFVGDGTVTVTIGISKGAMIKPNMATMLGYIATDAKVAQVQLDRLDAANKSFN  
RITIDGDTSTNDCMLIATGRAALPEVTOASGALFAALQAVLEVSMELAQAVRQGE  
GATKFTVGVNGGATHQECIDVGYAVAHSPLIKTAIFASDPNNGRILAAVGRAGVANL  
DYSKIDVFDGDCVIAISGRGAASTTEOGNAVVAQAIEGIRIELGRCTCSETTWTDDL  
SHEYKINAEIYRT  
complement(1455. .4205)  
/gene="seca"  
/locus\_tag="PA4403"  
complement(1455. .4205)  
/gene="seca"  
/locus\_tag="PA4403"  
/note="Protein name confidence: Class 2 (high similarity to functionally studied protein)"  
/codon\_start=1  
/transl\_table=11  
/product="secretion protein SecA"  
/protein\_id="AAG07791.1"  
/db\_xref="GI:9950635"  
/translation="MFAPLLKFLGSKNERDVKRMARVQAQINALEPQMVALSDEQLK  
AKTAEQRYAKGETLDLLPEAFVAVVREAGKRVGMGRHFDVOLIGMTLHDKGKIAEM  
RTGEGKTLVGLTPVYLNALSGKGVHVTVNDYLARDANMRPLVIGELGSGVVPF  
QPEDKRAAYAADITGTTNNEFGDYLRDNMAFLDDKFORELNFVAVYRDSILIDE  
ARTPLIISGQAEDESSSELYIKINKLIPRLKRVVEEGKPTTEEGHYSIDEKTRVELNE  
QGHQFIEDLISQNLGEGESLYSAHNLSLLTHVYAAALRAHTLFRHVEYIVGQDQIL  
LIDHTGRTMPGRRLSEGLHQAPEAKEGPLQAESQTLASTTQNTFQNTFRLYNKLGWGTG  
TADTEAFERQIYGLDVVITPRTARKDNLDVLTQEKVAAITDITKQCAQLGR  
PILVGPASIESSEYVSKLQEGIEHKVLNAKYHEKEETIAQAGAGSVTIATNAG  
RGTDILGNWEVEVALENPTTEQIAQIAEKWKHQHQVIEAGGLHVIASERHESRR  
IDNLRAGRAGQDPSRPFYLSLDNLMRIFASDRVKNFMKALGMSQSGAIBHMYT  
NIAEKARVVEGRNFDIRKOLLEDDVANEQRKVIYHMRNTLLSAEDVGETIEFRREE  
TISATNQHPQPSLQEQMDIEGLEAALYSDFAVRLPIQOMLDEDKLYETETRSKIL  
EOTVAAVYEKEELAGAEARAFKQMLRLVDLDDHLSTMDHLRHGILHRCYAKN  
PQOEYRESFTLQELLDSFKRDTIRVLSHVQVRREDPAEEEARLREAEELAKRMQF  
QHRAESMQAVAGEEELPEGPAPVVPLEPVRNEQKIGRNEKPCGSGKKYKHCHGQ  
LD  
complement(4339. .5262)  
/locus\_tag="PA4404"  
complement(4339. .5262)  
/locus\_tag="PA4404"  
/note="Protein name confidence: Class 4 (homologs of previously reported genes of unknown function, or no similarity to any previously reported sequences)"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAG07792.1"  
/db\_xref="GI:9950636"  
/translation="MHIIFLSRHHGAARSVHLQPRLLMLAGACVVGASLLAGGFMGAM  
LFRAPAPVAGIEVDAAALQAASAEISSAREDAQMDALSIVHLELQARLVRLDALGER  
LTELADVDPEDFSLVSGGGVDEPLEGSAYAPPFMATLDLEAARLDTREQOLEVL  
EOLLADKHISQSTYLAGQPVROGYISSPFRVRNVPVSGRLSMHKGIDFAAPAGSDVVA  
VAGVYVWTSGRKNGYCNVVEVGHADCTTLYAHNQKNLVEVGLDVRKQVLAQVSGTG  
RSTGYVHVHEVMKDGVRVMPQSFARASVE"  
5409. .5804  
/locus\_tag="PA4405"  
5409. .5804  
/locus\_tag="PA4405"  
/note="Protein name confidence: Class 4 (homologs of previously reported genes of unknown function, or no similarity to any previously reported sequences)"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAG07793.1"  
/db\_xref="GI:9950637"  
/translation="MLFAEQRICRLOQLLESQLOPAARPYCHVASWRGCGCLLIYTD  
GHWATRLYQRRRLQRLQAEFEFANLAKILLFKVQPTTSTQRAARSLTUSAVAENL  
RESAEIVGNPRLREALERLASRAKRSQE"  
complement(5865. .6776)  
/gene="lpxC"  
/locus\_tag="PA4406"  
/note="synonyms: envA, asmB"

## gene



CDS complement(5865. .6776)  
/gene="lpxC"  
/locus\_tag="PA4406"  
/note="Protein name confidence: Class 1 (function experimentally demonstrated in P. aeruginosa); Alt protein name: EnvA protein"  
/codon\_start=1  
/transl\_table=11  
/product="UDP-3-O-acyl-N-acetylglucosamine deacetylase"  
/protein\_id="AA07794.1"  
/db\_xref="GI:9950638"  
/translation="MIKQRLKNIIRATGVGLHSGEKVYLLKLPAPVDVTGIVFCRTDL DPVVEIPARENVGETMTSTLVKGDVDTVEHLLSAMAGLGIDNAYVELSASEVPI MDGSAFVLIQAGQEQAARFKIRIKREVSVEGDKRAVFPVDFGKFSVEIDF DHPVFRGTQASVDFSTSFVKVSRARTFGFMRDIEYLRSONLALGGSVENAIVVD ENRVLNEDGLRYEDEFVKHKILDAIGDLYLLGSLNLI GEFGRFKSGHALNQLLRLTIA DKDAWEVTFEDARTAPISVMPAAV"  
gene complement(6889. .8073)  
/gene="ftsZ"  
/locus\_tag="PA4407"  
/locus\_tag="PA4407"  
complement(6889. .8073)  
/gene="ftsZ"  
/locus\_tag="PA4407"  
/note="Protein name confidence: Class 2 (high similarity to functionally studied protein)"  
/codon\_start=1  
/transl\_table=11  
/product="cell division protein FtsZ"  
/protein\_id="AA07795.1"  
/db\_xref="GI:9950639"  
/translation="MFELVDNIQTAVIKVIGVGGGGNVAHVMAKNNVEGVEFICAN TDAQLNIAARTVQLPGVTKGLGAGANPEVGRQALEDREIRISELEGADVFIT TGMGGGTGAAPLIAVAKEMGILTVAVTPFPFEGRKRMQIADGIRALAESVDS LITPNEKLITILGKDSLAAFAKADVDVLAVAGVRSIDIKRPGMINVDPAVKTVM SEMGMAMTGACSSPNRAREATEAARNPDLLEDVNLQAGRIILVNITAGPDLISGEY SDVGNILEQFASERHATVKGTVIDAMRDELHVTIVATGLGARLEKPKVVDVNTVQGS AAQAAFAQREQQSVNRLDRLPTVMRNQSHGSAATAKUNLPQDDLDLYLPAFLRQ AD"  
gene complement(8124. .9377)  
/gene="ftsA"  
/locus\_tag="PA4408"  
Query Match 99.9%; Score 1449; DB 1; Length 24000;  
Best Local Similarity 100.0%; Pred. No. 2.6e-163;  
Matches 1449; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2 GTGCTGATCGCCTCGCCACCTTGAAGCTGCTTGAAGACGAGACGATGAGCCTCAT 61  
DB 16314 GTGCTGATCGCCTCGCCACCTTGAAGCTGCTTGAAGACGAGACGATGAGCCTCAT 16255  
QY 62 CGCCTCGACCACTTCGCGATCGTTGTGGCCTCGGCAAGAGCGCATGTCCTTGGTGGC 121  
DB 16254 CGCCTCGACCACTTCGCGATCGTTGTGGCCTCGGCAAGAGCGCATGTCCTTGGTGGC 16195  
QY 122 CTACCTGCGCGCGCGGCTTGCCTTTCGCGTGGTGATACCCGAGAACCCCGCGGA 181  
DB 16194 CTACCTGCGCGCGCGGCTTGCCTTTCGCGTGGTGATACCCGAGAACCCCGCGGA 16135  
QY 182 GCTGGCCACCTGGTCCCAAGTATCCGAGGTGAAGTGGTTCGGCGCAACTCGACGC 241  
DB 16134 GCTGGCCACCTGGTCCCAAGTATCCGAGGTGAAGTGGTTCGGCGCAACTCGACGC 16075  
QY 242 CGAGTTCTCTGCTCCGCGCCGCAACTCTATGTACGCCCGCGCTTGTGCTGCGCACCC 301  
DB 16074 CGAGTTCTCTGCTCCGCGCCGCAACTCTATGTACGCCCGCGCTTGTGCTGCGCACCC 16015  
QY 302 TGCCTGGTACAGCCCGCGCGAAGCGTGCATCTCCGGTGACATCGATCTCTTTCGC 361  
DB 16014 TGCCTGGTACAGCCCGCGCGAAGCGTGCATCTCCGGTGACATCGATCTCTTTCGC 15955  
QY 362 CCGCGAGCGAAGCCCGCGATCGTCGCATCACCAGTTCACCGGTCGAACCGGACCCGTCAC 421  
DB 15954 CCGCGAGCGAAGCCCGCGATCGTCGCATCACCAGTTCGAACCGGTCGAACCCGTCAC 15895

QY 422 CACCTTGTGGCGAAATGGCGGTGGCGCGACAAAGCGTGTGCGCGTGGCGGCAACT 481  
DB 15894 CACCTTGTGGCGAAATGGCGGTGGCGCGACAAAGCGTGTGCGCGTGGCGGCAACT 15835  
QY 482 CGGCACCCCGCGGTGCGACCTGCTGGCGGACGACATCGAGCTGTACGTGTGGAGCTGTC 541  
DB 15834 CGGCACCCCGCGGTGCGACCTGCTGGCGGACGACATCGAGCTGTACGTGTGGAGCTGTC 15775  
QY 542 GAGCTTCCAGCTGGAACCTGGATCGCTCAACGCCAGGTGGGACCGTGTCTGAAAGCT 601  
DB 15774 GAGCTTCCAGCTGGAACCTGGATCGCTCAACGCCAGGTGGGACCGTGTCTGAAAGCT 15715  
QY 602 CAGGGAAGACCATATGATCGCTACGACGGATGGCTGACTACACCTGGGCAAGACACCG 661  
DB 15714 CAGGGAAGACCATATGATCGCTACGACGGATGGCTGACTACACCTGGGCAAGACACCG 15655  
QY 662 GATCTTCCGCGTGGCGCCCGAGGTGGTGAATCGCGCCGATGCGCTGACCCGACCGCT 721  
DB 15654 GATCTTCCGCGTGGCGCCCGAGGTGGTGAATCGCGCCGATGCGCTGACCCGACCGCT 15595  
QY 722 GATCGCGATACCGTGGCGTGGCTGGCTGGCTGAGAGTGGCTGGCTGAGCAAGCGGACTTCAAGGCCTT 781  
DB 15594 GATCGCGATACCGTGGCGTGGCTGGCTGGCTGAGAGTGGCTGGCTGAGCAAGCGGACTTCAAGGCCTT 15535  
QY 782 CGGCTGATCGAGGAAGACGGCCAGAGTGGCTGGCTGGCTGAGCAAGCGGCTGGCTGGCTGGC 841  
DB 15534 CGGCTGATCGAGGAAGACGGCCAGAGTGGCTGGCTGGCTGAGCAAGCGGCTGGCTGGC 15475  
QY 842 GGTGGGCACTGAAGATCCGTGGCGCCACAACTATTTCACAGGCGTGGCGCGCTGGC 901  
DB 15474 GGTGGGCACTGAAGATCCGTGGCGCCACAACTATTTCACAGGCGTGGCGCGCTGGC 15415  
QY 902 GCTGGCCATCGGTGCGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGC 961  
DB 15414 GCTGGCCATCGGTGCGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGC 15355  
QY 962 CGGCTTGGCTCATCGCTGCCAGTGGGTACGGAGCGGAGCGGCGTGGCTGGCTGGCTGGCTGGC 1021  
DB 15354 CGGCTTGGCTCATCGCTGCCAGTGGGTACGGAGCGGAGCGGCGTGGCTGGCTGGCTGGC 15295  
QY 1022 TTCAGGCGCAACCAACGTCGGCGCGCTGGCGCGGATCGAGGGGCTGGGTGGCGGAT 1081  
DB 15294 TTCAGGCGCAACCAACGTCGGCGCGCTGGCGCGGATCGAGGGGCTGGGTGGCGGAT 15235  
QY 1082 CGAGGCAAGCTGGTGTGCTGCGCGCGGAGAGCGGCAAGGCGCGGATTTCCATGACCT 1141  
DB 15234 CGAGGCAAGCTGGTGTGCTGCGCGCGGAGAGCGGCAAGGCGCGGATTTCCATGACCT 15175  
QY 1142 CGCGAGCGGTCGCGCGCTTCTGCGCGCGGCTGGTACTGCTTGGCGCTGAGCGCGGCT 1201  
DB 15174 CGCGAGCGGTCGCGCGCTTCTGCGCGCGGCTGGTACTGCTTGGCGCTGAGCGCGGCT 15115  
QY 1202 GATTGCCAGGACTGGGCAACCGGTTACCGGTGTGTCGCGCTGCGCAACGCTGGAGCAAGC 1261  
DB 15114 GATTGCCAGGACTGGGCAACCGGTTACCGGTGTGTCGCGCTGCGCAACGCTGGAGCAAGC 15055  
QY 1262 AGTCCGCGAGCGCGCGAGCTGGCCCGGAGGCGATGGGTGCTGTGTCGCGCGCTG 1321  
DB 15054 AGTCCGCGAGCGCGCGAGCTGGCCCGGAGGCGATGGGTGCTGTGTCGCGCGCTG 14995  
QY 1322 CGCGAGCTTGACATGTTCAAGAACTTCAAGAACTTCAAGAACTTCAAGAACTTCAAGAACTT 1381  
DB 14994 CGCGAGCTTGACATGTTCAAGAACTTCAAGAACTTCAAGAACTTCAAGAACTTCAAGAACTT 14935  
QY 1382 AGAGGAGCTAGCGTATGCTGCGTGTGTCGCGCTTCCGCTCGCGCTGCTGAGCGCG 1441  
DB 14934 AGAGGAGCTAGCGTATGCTGCGTGTGTCGCGCTTCCGCTCGCGCTGCTGAGCGCG 14875  
QY 1442 CACGGCATC 1450  
DB 14874 CACGGCATC 14866

```

RESULT 2
AY008276
LOCUS      5280 bp    DNA    linear    BCT 19-JUL-2001
DEFINITION MurD (murD) genes, complete cds.
ACCESSION AY008276
VERSION    1
KEYWORDS   MurD (murD) genes, complete cds.
SOURCE     AY008276.2 GI:12007715
  ORGANISM Pseudomonas aeruginosa
  Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
  Pseudomonadaceae; Pseudomonas.
  1 (bases 1 to 5280)
  Azzolina, B.A., Yuan, X., Anderson, M.S. and El-Sherbeini, M.
  The cell wall and cell division gene cluster in the Mra operon of
  Pseudomonas aeruginosa: cloning, production, and purification of
  active enzymes
  Protein Expr. Purif. 21 (3), 393-400 (2001)
  21178826
  11281713
  2 (bases 1 to 5280)
  El-Sherbeini, M. and Azzolina, B.
  Pseudomonas aeruginosa mray, murD, murE, and murF genes
  Unpublished
  3 (bases 1 to 5280)
  El-Sherbeini, M. and Azzolina, B.
  Direct Submission
  Submitted (20-SEP-2000) Biochemistry, Merck Research Laboratories,
  126 E. Lincoln Avenue, Rahway, NJ 07065, USA
  4 (bases 1 to 5280)
  El-Sherbeini, M. and Azzolina, B.
  Direct Submission
  Submitted (02-JAN-2001) Biochemistry, Merck Research Laboratories,
  126 E. Lincoln Avenue, Rahway, NJ 07065, USA
  Sequence update by submitter
  On Jan 2, 2001 this sequence version replaced gi:10719753.
FEATURES
     Location/Qualifiers
         1..5280
             /organism="Pseudomonas aeruginosa"
             /mol_type="genomic DNA"
             /db_xref="taxon:287"
         1..1464
             /gene="murE"
         1..1464
             /gene="murE"
         6..diaminopimelate ligase"
             /note="UDP-N-acetylmuramoyl-L-alanyl-D-glutamate:meso-2,
             /codon_start=1
             /transl_table=11
             /product="MurE"
             /protein_id="AAG45235.1"
             /db_xref="GI:12007716"
             /translation="WPMISQLFPOAERDLIRELTDLSHGVRPVDLFLTPVGGHODG
             RAHIALDTKGTATVAYEAGAGELPPSDAPLIYVKGIAQLSAVAGRYPPGGRLD
             LIGVTNGKTSVQLVAQDLGRCIGVTLGTGFGYGALESRRHTPPDPLVAQAT
             LATLQAGARVAVSSHGLDQGVAAFGDIYFTNLSDRLDYLHGSMEAYAAKA
             KLFAPDPLRCRVINLDDGRRLAGEEQDSELITYSLTSSAFLYCRASPASPTAR
             RSSIAKOGLLRSLPLGRNLNLLAAVGLIGVPLGDIILRTLPOLQGPVGRMORL
             GGSGPLVVVDYVHTPDLEKLEALRPHAAALLLFCGCGDRDAGKRPMLAATER
             LADELIVTDNPTASAAIADIRKGFNAADKVTFLPSRGEAIAHLIASAAVDVVVL
             LAGKHEDIQEIDGVRHPFSDIEQAERAAWVEVPHA"
             1457..2833
             /gene="murF"
             1457..2833
             /gene="murF"
             /note="recombinant enzyme catalyzed the ATP-dependent
             addition of D-alanine-D-alanine to the
             UDP-N-acetylmuramyl-L-alanine-D-glutamate-m-Dap precursor"
             /codon_start=1
             /transl_table=11
             /product="MurF"
             /protein_id="AAG45236.1"

```

```

/db_xref="GI:12007717"
/translation="MLEPLRLSOLTVALDARLIGEDAVFSAVSTDSRAITPGOLFIAL
SPRPDGHYDLAEVAAKGAVALVEREADAPLPOLLVRDTRAALGLGLNRKKEFG
PLAMFTSGSKTAVKEMLASILTKQAGDAESVLATRGNLNNDLGVPLTLQLAPQHS
AVIEGASGKELIAYVELTRPHVAILTINAGTAHVGEFGPEKIDRAKIEGLEGLAAD
GTAVLNLDKADFDTWKARASGRPLLTFTSLDRFPQADFPRAADLQKRAVAGFRLQGVAG
EAQVOLNLGIRHNVAALAAAAAALGVPLDGIIVAGLQALQPKVRAVAQLTASGLR
VIDDSYNANPASMLAAILDLSGFSRTVLVLDGMDGLSGMAEQAREVGAYAAKYSVA
LYAVGPLMAHVAQFATGCRHFADQASLIGALATEQPTTILIKSRSNAAMDKVVAAL
CQSSSESH"
2833..3915
/gene="mray"
2833..3915
/gene="mray"
/note="phospho-N-acetylmuramoyl-pentapeptide translocase"
/codon_start=1
/transl_table=11
/product="MurF"
/protein_id="AAG22121.1"
/db_xref="GI:10719754"
/translation="MLLLAEYLQOFYKGFVQFYLTLRGLSVLTALSILSLWLGPMW
IRTLQIPIQOAVRNDGPOSHLSKKGKPTPMGGALILTAISTLLWADLSNRVYVVL
VVTLLGCAIGWDDYRKVKIEKNSRGLPSRWKFWQSVFGIGAAYFLYMTAEPIETTL
IVPMKLSVEIQIGFIFFVLITYFVIGSSNAVNLTDGLDLAIMPVTVVAGALGICIL
SGNVKFAEYLLIPNVPGAGELIVFCAALYVAGLGFELWNTYPAQVPMGVDVGALALGA
LGTIAIVVREIQLVFLTMGVFVNETLSVMIOVASFKLTRRRVFRMAPIHHHFLKGPW
DPRIVRFRWIIITVILVLIGLATLKLK"
3929..5275
/gene="murD"
3929..5275
/gene="murD"
/note="UDP-N-acetylmuramyl-L-alanine:D-glutamate ligase;
recombinant enzyme catalyzed the ATP-dependent
addition of D-glutamate to the precursor sugar peptide"
/codon_start=1
/transl_table=11
/product="MurD"
/protein_id="AAG45237.1"
/db_xref="GI:12007718"
/translation="MSLIAADHFRIVVGKSGMSLVRYLARRGLPFAVVDTRNPPE
LATLRAQYQVEARGELDAEFLCSARELYVSPGLSLRTPALVQAQAKVRISGDLDL
FAREAKAPIVATGSKNAKSTVTLTGEMAVADKRVAVGNGLTGTPALDLLADIELLY
LELSSFLQEDRLNNAEVATLVNSDEHMDRDMADYHLAKHRLFRGARQVYVNRAD
ALTRPLIADTVPCVFNKPFKAFGLIEEDQKWLAFQKGLKLPVGEKIRGAHNY
SNALAAIGHAVGLPFDAMLGALKAFLSGHRCQWRERQGVSYVDDSKATNVGCAAL
AAIEGIGADIDGKLVLLAGDGKADFDHDLREPVARFCRAVLLGRDAGLIAQAALNA
VPLVRVATLDEAVRQAALAREGDVALLSPACASLDMFKNFEERGLFAKAVEELA"
BASE COUNT      817 a 1781 c 1759 g 923 t
ORIGIN
Query Match      96.3%; Score 1396.2; DB 1; Length 5280;
Best Local Similarity 99.8%; Pred. No. 6.9e-157;
Matches 1398; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY      2 GTGCTGATCGGCTCGCCACCTTGAGCTCGGTGAGGACGAGAGAGATGAGCCTGAT 61
DB      3880 GTGCTGATCGGCTCGCCACCTTGAGCTCGGTGAGGACGAGAGATGAGCCTGAT 9399
QY      62 CGCCTCCGACCACTTCGCGCATCGTTGTCGCGCTCGGCAAGAGCGGCATGTCCCTGCTGGC 121
DB      3940 CGCCTCCGACCACTTCGCGCATCGTTGTCGCGCTCGGCAAGAGCGGCATGTCCCTGCTGGC 3999
QY      122 CTACCTGGCGCGCGCGGCTTCCTTTCCCGCTGTCGATACCCGAGAGAACCCGCCGA 181
DB      4000 CTACCTGGCGCGCGCGGCTTCCTTTCCCGCTGTCGATACCCGAGAGAACCCGCCGA 4059
QY      182 GCTGGCCACCTTCGCTGCCAGTATCCGAGGTGGAGTGGTGGCGGGAAGTCCGACGC 241
DB      4060 GCTGGCTACCTTCGCTGCCAGTATCCGAGGTGGAGTGGTGGCGGGAAGTCCGACGC 4119
QY      242 CGAGTTCTCTCTCGTCCGCCCGCGGAAGTCTATGTCAGCCCGCGGCTTGTCTGCTGGCACCCC 301
DB      4120 CGAGTTCTCTCTCGTCCGCCCGCGGAAGTCTATGTCAGCCCGCGGCTTGTCTGCTGGCACCCC 4179

```

QY	302	TCGCGTGTGTACAGCGCCGCGGAAAGGGGTGCGCATCTCCGGTGACATCGATCTCTTGGC	1382	AGAGGAGTAGCGTGATGCTG 1402
Db	4180	TGCGCTGTGTACAGCGCCGCGGAAAGGGGTGCGCATCTCCGGTGACATCGATCTCTTGGC	5260	ACAGGAGTAGCGTGATGCTG 5280
QY	362	CGCGAGGCGAAGGCCCGGATGTCGCCATACCGGTTCCACAGCGAAGACCGTGCAC	RESULT 3	
Db	4240	CGCGAGGCGAAGGCCCGGATGTCGCCATACCGGTTCCACAGCGAAGACCGTGCAC	AE016779	
QY	422	CACCTGTGTGGCGAAATGGCGGTGCGCGGACAAAGCGTGTGCGCGCGGCAACCT	DEFINITION	Pseudomonas putida KT2440 section 6 of 21 of the complete genome.
Db	4300	CACCTGTGTGGCGAAATGGCGGTGCGCGGACAAAGCGTGTGCGCGCGGCAACCT	ACCESSION	AE016779.1
QY	482	CGGCACCCCGCGCTGCGACCTGCTGCGCGGACGACATCGAGCTGTGCTGTTGGAGCTGTC	VERSION	AE016779.1
Db	4360	CGGCACCCCGCGCTGCGACCTGCTGCGCGGACGACATCGAGCTGTGCTGTTGGAGCTGTC	KEYWORDS	GI:26557023
QY	542	GAGCTTCCAGCTGGAAACCTCGGATCGGCTCAACCGCGAGGTGGCGACCGTGAACGT	SOURCE	Pseudomonas putida KT2440
Db	4420	GAGCTTCCAGCTGGAAACCTCGGATCGGCTCAACCGCGAGGTGGCGACCGTGAACGT	ORGANISM	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
QY	602	CAGCAAGACCATATGGATCGCTACGACGGCATGCTGACTACCACTGGCCAAAGCACCG	REFERENCE	1 (bases 1 to 30195)
Db	4480	CAGCAAGACCATATGGATCGCTACGACGGCATGCTGACTACCACTGGCCAAAGCACCG	AUTHORS	Nelson,K., Paulsen,I., Weinel,C., Dodson,R., Hilbert,H., Fouts,D., Gill,S., Pop,M., Martins Dos Santos,V., Holmes,M., Brinkac,L., Beanan,M., DeBoy,R., Daugherty,S., Kolonay,J., Madupu,R., Nelson,W., White,O., Peterson,J., Khouri,H., Hance,I., Lee,P., Holtzapple,E., Scanlan,D., Tran,K., Moazzez,A., Utterback,T., Rizzo,M., Lee,K., Kosack,D., Moesti,D., Wedler,H., Lauber,J., Hoheisel,J., Straetzl,M., Helm,S., Kiewitz,C., Eisen,J., Timmis,K., Duesterhoft,A., Tummier,B. and Fraser.C.
QY	662	GATCTTCCGCGTCCGCCAGTGTGATCGCGCGATGCCCTGACCCGACCGCT	TITLE	Complete genome sequence and comparative analysis of the metabolically versatile Pseudomonas putida KT2440
Db	4540	GATCTTCCGCGTCCGCCAGTGTGATCGCGCGATGCCCTGACCCGACCGCT	JOURNAL	Environ. Microbiol. 4 (12), 799-808 (2002)
QY	722	GATCCCGATACCGTGGCTGTGCTGTGGCTGAAACAGCGCGACTTCAAGCGTTT	REFERENCE	2 (bases 1 to 30195)
Db	4600	GATCCCGATACCGTGGCTGTGCTGTGGCTGAAACAGCGCGACTTCAAGCGTTT	AUTHORS	Nelson,K., Paulsen,I., Weinel,C., Dodson,R., Hilbert,H., Fouts,D., Gill,S., Pop,M., Martins Dos Santos,V., Holmes,M., Brinkac,L., Beanan,M., DeBoy,R., Daugherty,S., Kolonay,J., Madupu,R., Nelson,W., White,O., Peterson,J., Khouri,H., Hance,I., Lee,P., Holtzapple,E., Scanlan,D., Tran,K., Moazzez,A., Utterback,T., Rizzo,M., Lee,K., Kosack,D., Moesti,D., Wedler,H., Lauber,J., Hoheisel,J., Straetzl,M., Helm,S., Kiewitz,C., Eisen,J., Timmis,K., Duesterhoft,A., Tummier,B. and Fraser.C.
QY	782	CGGCTGTATCGAGGAGGCGCAGAGTGGCTGCGTTCCAGTTTCGACAGCTGCTGCC	TITLE	Direct Submission
Db	4660	CGGCTGTATCGAGGAGGCGCAGAGTGGCTGCGTTCCAGTTTCGACAGCTGCTGCC	JOURNAL	Submitted (05-NOV-2002) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
QY	842	GTTTGGCAACTGAAGATCCGTGGCGCCACAACTATTCCACGCGCTCGCGCGCTGGC	FEATURES	Location/Qualifiers
Db	4720	GTTTGGCAACTGAAGATCCGTGGCGCCACAACTATTCCACGCGCTCGCGCGCTGGC	source	1. 30195
QY	902	GCTGGCCATGCGGTGCGCTGCTGCGCCATGCTGCGCGCGCTGAAGCGCTTTTC	gene	/organism="Pseudomonas putida KT2440"
Db	4780	GCTGGCCATGCGGTGCGCTGCTGCGCCATGCTGCGCGCGCTGAAGCGCTTTTC	CDS	/mol_type="genomic DNA"
QY	962	CGGCTGTATCGAGGAGGCGCAGAGTGGCTGCGTTCCAGTTTCGACAGCTGCTGCC		/strain="KT2440"
Db	4840	CGGCTGTATCGAGGAGGCGCAGAGTGGCTGCGTTCCAGTTTCGACAGCTGCTGCC		/db_xref="taxon:160488"
QY	1022	TTCCAAGGCCAACACGTGCGCGCCCGCTGCGCGCGATCGAGGGCGTGGTGCAGACAT		89..712
Db	4900	TTCCAAGGCCAACACGTGCGCGCCCGCTGCGCGCGATCGAGGGCGTGGTGCAGACAT		/gene="ssspA"
QY	1082	CGACGGCAAGCTGTGCTGCTGCGCGGAGAGCGGCAAGCGCGCGATTTCCATGACCT		89..712
Db	4960	CGACGGCAAGCTGTGCTGCTGCGCGGAGAGCGGCAAGCGCGCGATTTCCATGACCT		/locus_tag="pp1320"
QY	1142	GCGGAGCGGTCGCGCGCTTCTGCGCGCGTGTGCTGCTGCGCGTGAAGCGCGGCT		/note="similar to GP:9864180; identified by sequence similarity; putative"
Db	5020	GCGGAGCGGTCGCGCGCTTCTGCGCGCGTGTGCTGCTGCGCGTGAAGCGCGGCT		/codon_start=1
QY	1202	GATTGCCCAAGGCACTGGGCAACCGGCTGCTGCGCGCTGCGCAAGCGCTGGAGGAGC		/transl_table=11
Db	5080	GATTGCCCAAGGCACTGGGCAACCGGCTGCTGCGCGCTGCGCAAGCGCTGGAGGAGC		/product="stringent starvation protein A"
QY	1262	AGTCCGGAGCGCGCGAGCTGGCGGAGAGCGGATGCGGTGCTGCGCGCGGCTG		/protein_id="AAN66944.1"
Db	5140	AGTCCGGAGCGCGCGAGCTGGCGGAGAGCGGATGCGGTGCTGCGCGCGGCTG		/db_xref="GI:24982777"
QY	1322	CGCGAGCTGGACATGTTCAAGAACTTCGAAGAACCGGACCGCTGTCGCGCAAGCGCT		/translation="MGATNRLACYSDPADHYSHRVLRLAEGVSVQLIDVDPAHLPR KLAENYPYGSVETLVDRLALYETVYMEYLEEYPPHPLMPVYPVARGNSRLMHRI QRDWCALADTFLDPRSEAAETARKALRESLTGVSPLSFSEAFCSMDRSQSLVDCCLL PTLWLPLVIGIELPQAKPLLDYMERQFAREFQASLSVSVERMKL"
Db	5200	CGCGAGCTGGACATGTTCAAGAACTTCGAAGAACCGGACCGCTGTCGCGCAAGCGCT		693..1154
				/gene="ssspB"
				/locus_tag="pp1321"
				693..1154
				/gene="ssspB"
				/locus_tag="pp1321"
				/note="similar to GB:S50157, GB:X60221, SP:P24539, and PTD:509291; identified by sequence similarity; putative"
				/codon_start=1
				/transl_table=11









Qy	21	CGTTGAAGCTGCGTTGAGGACGAGAGAGATGAGCCTGATCGCTCCGACCACTTCCGC	80
Db	291882	CGTTGCCACACTGAAACTGAGGTAGAACGAGTGTCCCTGATCGTTTCCGACCGCTTCCGC	291823
Qy	81	ATCTGTGTGCGGCTTCGCAAGAGCGGCATGTCCCTGTGCGCTACCTTGGCGCGCCGCGC	140
Db	291822	ATTGTTGTGGGCTTGGCAAGAGTGGCATGTCCCTGTGCTTCCCTGGCGAACC6GGGC	291763
Qy	141	TTGCGCTTTCGCGTGGTGTGATACCCGAGAGAAACCCGCGGAGCTGGCCACCTCGCTGCC	200
Db	291762	GTGTCCTTTGCTTGGCGGACACGCGGAAATCTCTGAGCTGGGCGACTTTCGCTGCT	291703
Qy	201	CAGTATCCGCGAGGTGGAAGTGGTTCGGCGGAACTCGACGCGGAGTTCCTCTGCTGCGCC	260
Db	291702	GACTATCTCAGTGGAAAGTGGGTGTGGGAACTTGGACGTGGACTTCTCTGCGCTGCC	291643
Qy	261	CGGAACTCTATGTACGCGCGGCTTGTGCTGTGGCGACCCCTGCGCTGTGTACAGCGCGCC	320
Db	291642	GACGAGCTTTATGTACGCGCTGCTGTGCGCTTGGCCACACCGCGCTTTCGACAGCGCGAT	291583
Qy	321	CGGAAGGCGTGGCATCTCCGCTGACATCGATCTTTCGCGCGGAGGCGAAGCGCCCG	380
Db	291582	GCACGCGCGCAAGCTGTCCGCGGACATCGAACTGTTTTCGCGCTTACGCCAAGGCGACCG	291523
Qy	381	ATCTGCGCATCACCGGTTCCAAACGCGAAGAGCACCGTGACACCCCTGGTGGGCGAAATG	440
Db	291522	GTGATTGCGATTACCGGTTTCCAAACGCGAAGAGCACCGTGACACCGTGTGCGCGAGATG	291463
Qy	441	GCGGTGGCGCGGACAGCGTGTCCGCTGGCGGCAACCTTCGGACACCCCGCGCGCTGCAC	500
Db	291462	GCGCAGCTGCGCGGAAGCGTGTCTCTGTGGCGGCAACCTTGGCACCCCGCGCTGGAC	291403
Qy	501	CTGCTGGCGACGACATCGAGCTGTACGTGTGTGAGCTGTGAGCTTCCAGCTTGGAAACC	560
Db	291402	CTGCTCAGGATGATGTGCGATGTGTACGTGTGATGAAATGTCCAGTTTCACTCGAAACC	291343
Qy	561	TGCGATTCGCTCAACCGCGAGTGGCGACCGTGTGTAACGCTCAGCGAAGACCATATGAT	620
Db	291342	ACTGATCAGCTCAATGCGGAAGTGGCGACCGTGTGTAACATCAGCGAAGACCATATGAC	291283
Qy	621	CGCTAGCGGCGATGCTGACTTACCACTGGCAAGCAGCGATCTTCCGCGGTGCCCGC	680
Db	291282	CGTTATTCGCTGCTCGCGGCTTATCACCTGCGCAAGCACCGTATTTTCCGTGGCGCGGG	291223
Qy	681	CAGTCTGTGGTGAATCGGCGCGATGCCCTGACCCGACCGCTGATCGCGCATACCGTCCG	740
Db	291222	CAAGTAGTGTCAATCGCGAGGACGCGCTGTCCGCTCGCTGATCGCGGAGGCGCTGCT	291163
Qy	741	TGCTGTGCTTTCGCGCTGACAAAGCGGAGCTTCAAGGCTTTCGCGCTGATCGAGAGAC	800
Db	291162	TGCTGGAGCTTTCGCGCTTCAACAGGCTGACTTCCACGCGTGTGCTGCTGTAAGAAC	291103
Qy	801	GCCAGAGTGTGCTGCGCTTCCAGTTCGACAAAGCTGCTGCGGTTGGCGAACTCAAGATC	860
Db	291102	GGCGAAATTCCTGGCATTCAGTTCGAGAACCTGATGCGCGGTGAGCGAGCTGAAAGTT	291043
Qy	861	CGTGGCGCCACAACATTTTCCAAACGCGTTCGCGCGCTGGCGCTGGGCGCATCGCTCGC	920
Db	291042	CGCGGGCACAAATCAGCTTAACGACTGGCAGCCCTGGCGCTGGGCGATCGCTGGCG	290983
Qy	921	CTGCGCTTCGACCGCATGCTCGCGCGGTGAAGCGGTTTTCGCGCTGGCTCATCGCTGC	980
Db	290982	CTGCGCTTCGACCGCATGCTTTCAGCGCTGCGTGAAGTTTTCACGCGCTTGGACCGCTGC	290923
Qy	981	CAGTGGGTACGCGGAGGCGGCGTGGCTGAGCTACTACGACGATTCACGCGCCACCAAGTC	1040
Db	290922	CAATGGCTGCGTGAGCGCAATGGCGTGGACTACTACAACGATTCACAGGCGCCACCAAGCTG	290863

Query Match	57.18;	Score 827.8;	DB 1;	Length 311600;	
Best Local Similarity	73.68;	Pred. No. 6.5e-90;			
Matches 1052;	Conservative	0;	Mismatches 378;	Indels 0;	Gaps 0;

AL627265	251050 bp	DNA	linear	BCT 06-JUN-2002
Salmonella enterica serovar Typhi (Salmonella typhi) strain Ctr18,				
complete chromosome; segment 1/20.				
AL627265	AL513382			
AL627265.1	GI:16501283			
Salmonella enterica subsp. enterica serovar Typhi				
Salmonella enterica subsp. enterica serovar Typhi				
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;				
Enterobacteriaceae; Salmonella.				
1 (bases 1 to 251050)				
Parkhill, J., Dougan, G., James, K.D., Thomson, N.R., Pickard, D.,				
Wain, J., Churcher, C., Mungall, K.L., Bentley, S.D., Holden, M.T.G.,				
Sebahia, M., Baker, S., Basham, D., Brooks, K., Chillingworth, T.,				
Connor, P., Cronin, A., Davis, P., Davies, R.M., Dowd, L., White, N.,				
Farrar, J., Feltwell, T., Hamlin, N., Haque, A., Hien, T.T., Holroyd, S.,				
Jagels, K., Krogh, A., Larsen, T.S., Leather, S., Moule, S., O'Gaora, P.,				
Parry, C., Quail, M., Rutherford, K., Simmonds, M., Skelton, J.,				
Stevens, K., Whitehead, S. and Barrrell, B.G.				
Complete genome sequence of a multiple drug resistant Salmonella				
enterica serovar Typhi Ctr18				
Nature 413 (6858), 848-852 (2001)				
21534947				
11677608				
2 (bases 1 to 251050)				
Parkhill, J.				
Direct Submission				
Submitted (25-OCT-2001) Submitted on behalf of the Salmonella				
sequencing team, Sanger Centre, Wellcome Trust Genome Campus,				
Hinxton, Cambridge CB10 1SA, UK				
E-mail: parkhill@sanger.ac.uk				

Notes:

Details of S. typhi sequencing at the Sanger Centre are available on the World Wide Web.  
(URL, [http://www.sanger.ac.uk/Projects/S\\_typhi/](http://www.sanger.ac.uk/Projects/S_typhi/)).

FEATURES	Location/Qualifiers
source	1. 251050
	/organism="Salmonella enterica subsp. enterica serovar Typhi"
	/mol_type="genomic DNA"
	/strain="Ctr18"
	/db_xref="taxon:90370"

```

gene
190. .255
/genes="STY0001"
/notes="synonym: thrL"
190. .255
/genes="STY0001"
/notes="Orthologue of E. coli thrL (LPT_ECOLI); Fasta hit
to LPT_ECOLI (21 aa), 86% identity in 21 aa overlap"
/codon_start=1
/transl_table=11
/product="thr operon leader peptide"
/protein_id="CAD01154.1"
/db_xref="GI:16501284"
/db_xref="SPTREMBL:O8XG12"
/translation="MNRISTTTTITTTTNGAG"
337. .2799
/genes="STY0002"
/notes="synonym: thrA"
337. .2799
/genes="STY0002"
/notes="Orthologue of E. coli thrA (AKH_ECOLI); Fasta hit
to AKH_ECOLI (820 aa), 94% identity in 820 aa overlap"
/codon_start=1
/transl_table=11
/product="aspartokinase I/homoserine dehydrogenase I"
/protein_id="CAD01155.1"
/db_xref="GI:16501285"
/db_xref="SPTREMBL:Q8Z9R7"
/translation="NRVLKFGCTSVANAERFLRVADILLESNSRQGOVATVLSAPAKIT
NLHVMIEKTIGQDALNISDAERIFSDLLAGLASAQGFPLARLKNVVEEFAQIK
HVLGSLGQCPSDINAALICRGENKMSIAIMAGLLEAGRHRTVIDPVEKLVAGHY
LSTVDIAESTRIAASQIPADHMLMAGTAGEKGEDELVILGRNGSDYSAVALAACF
RADCEIIVTDVGVTCDPQRPVDAARLLKMSYQAMLSYFCAKVLHPRTTTPAQL
QIPCLIKTNGPQAGTLLIGASDDNLPVKGISLNNMAMFVSQPGMKMGIMAAAR
VFAAMSAGISVVLITQSSEYSISFCVQPCARARMAQDEFEYLEKEGLELPLAV
TERLAIISVVGDMFTLGCISAKFAALARANINIVIAQGSSEISISVVVNDATT
GYRVHQLMFDQVIEVFVIGVGGVGGALLEGQLKQQLNKKHIDLRVCVANSKA
LUTNHGLNDWQLEAQAANFNLRLIRLVKEYLHLLNPVIVDTSSQVADQYAD
FIREGFVYTPNKKATSSMDYHQLRFAAQSRRKFLYDTNVGAGLPVIEQLNLLN
AGDELOKFSIGLSGSISFTFLEGMSSQATALAREMGYTEPDPDLSQMDVARK
LILARETGRELELSDIVLEVPDPEFDASGDVTAFAHLPOLDDAFAARVAKARDEG
KVLRYVGNTEEDGVCVKTAEDVGDNDPLFKVNGENALAFYSHYIQLPLVLRGYGAG
NDVTAAGVPADLLRLSLWLGV"
337. .1101
/genes="STY0002"
/notes="pfam match to entry PF00696 aakkinase, Amino acid
kinase family, score 207.70, E-value 1.7e-58"
343. .369
/genes="STY0002"
/notes="PS00324 Aspartokinase signature"
1819. .2781
/genes="STY0002"
/notes="pfam match to entry PF00742 Homoserine dh,
Homoserine dehydrogenase, score 535.60, E-value 3.4e-157"
2314. .2382
/genes="STY0002"
/notes="PS01042 Homoserine dehydrogenase signature"
2801. .3730
/genes="STY0003"
/notes="synonym: thrB"
2801. .3730
/genes="STY0003"
/notes="Orthologue of E. coli thrB (KHSE_ECOLI); Fasta hit
to KHSE_ECOLI (310 aa), 94% identity in 308 aa overlap"
/codon_start=1
/transl_table=11
/product="homoserine kinase"
/protein_id="CAD01156.1"
/db_xref="GI:16501286"
/db_xref="SWISS-PROT:Q8XGP5"
/translation="MVKYVAPASSAMSVGFDVGAAVTVPDGTLLGDVSVVEAADHF
RLHNGRADKLPPREINIVYQCWERFCQALGKIPVAMTLEKNMIGSGLGSSACS
VVAALVAMNEHCGRPLNDRLLALMELEGRISGSIHVDNVAPCPFLGGMGLIENGSI
ISQVPGFDENLWLAYPGKIVSTAEARAILPAQYRQDCTAHGRHLAGFIHACYRSR

```

```

misc_feature
3005. .3196
/genes="STY0003"
/notes="pfam match to entry PF00288 GHMP_kinases, GHMP
kinases putative ATP-binding proteins, score 79.20,
E-value 5.7e-21"
3068. .3103
/genes="STY0003"
/notes="PS00627 GHMP kinases putative ATP-binding domain"
3734. .5020
/genes="STY0004"
/notes="synonym: thrC"
3734. .5020
/genes="STY0004"
/notes="Orthologue of E. coli thrC (THRC_ECOLI); Fasta hit
to THRC_ECOLI (428 aa), 93% identity in 428 aa overlap"
/codon_start=1
/transl_table=11
/product="threonine synthase"
/protein_id="CAD01157.1"
/db_xref="GI:16501287"
/db_xref="SPTREMBL:Q8Z9R6"
/translation="MKLYNLKDHNEQVSFAQAVTQGLGKQGLFFPHDLPFSLFEID
ENLNQDFVRSARILSAGTIGDEIPQOILEERVAAFAFPAPVAQVEDYGCLELPHGP
TLAFKDFGFRFMAQMLTHISGDKPVTILTATSGDTCAAVAHAFYGLENVVLYLPRG
KISLQEKLFCTLGGNIETVADIGDFDAQALVKAQAFDDDEELKTALGNSANSINISR
LLAOLCYFEAVALPOGARNOLVISVPSGNGCDLTAGLAKSLGPKVRKFTAAANNIN
DYVPRLHDGKAPKATQNTLSNAMDVSQPNWPRVEELFRKIRWELFELGVAANDDT
TTPQRMRELKAGYISEPHAVAAYRALDQLPNGPEYGLFGLTGAHPAKFKESVILGE
TIALPEALAEARADPLLLSHLPAADFAALRKLMMTRQ"
3962. .4867
/genes="STY0004"
/notes="pfam match to entry PF00291 PALP,
pyridoxal-phosphate dependent enzyme, score 104.50,
E-value 2e-27"
4022. .4066
/genes="STY0004"
/notes="PS00165 Serine/threonine dehydratases
pyridoxal-phosphate attachment site"
complement(5114. .5887)
/genes="STY0005"
/notes="synonym: yaaA"
complement(5114. .5887)
/genes="STY0005"
/notes="Orthologue of E. coli yaaA (YAAA_ECOLI); Fasta hit
to YAAA_ECOLI (258 aa), 86% identity in 257 aa overlap"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAD01158.1"
/db_xref="GI:16501288"
/db_xref="SPTREMBL:Q8Z9R5"
/translation="WLIIISPAKTLDYQSPLATTRYTOPELLDHSQOLIQAROLSLAP
QISRLMGISDLADLNATRFHDWQPHFTPDNARQAILAFKGDVYITGLQRETFNDADF
FAQQHLRLMSGLYGLVRLDLMQPYRLENGIRLENPRGKDLQYFGFGDITDKLNALE
AQGRVYVNLASSEEYFKSPKKLDAELIKPVFLDEKNGKVKVSVFYAKKARGLSRF
IIEENLTKPEQLTAFDRGYFDEETSTQDELVFKRYEQ"
complement(5966. .7396)
/genes="STY0006"
/notes="synonym: yaaJ"
complement(5966. .7396)
/genes="STY0006"
/notes="Similar to Bacillus subtilis amino acid carrier
protein alst ALST_BACSU (Q45068; P40743) fasta
scores: E(): 0, 45.3% id in 477 aa, and to Alteromonas
haloplanktis na(+)-linked D-alanine glycine permease DGA
SW:DAGA_ALTHA (P30144) fasta scores: E(): 0, 36.5% id in
315 aa
Orthologue of E. coli yaaJ (YAAJ_ECOLI); Fasta hit to
YAAJ_ECOLI (476 aa), 76% identity in 476 aa overlap"
/codon_start=1
/transl_table=11

```



```

/product="putative amino-acid transport protein"
/protein_id="CAD01159.1"
/db_xref="GI:16501289"
/db_xref="SPTREMBL:Q8Z9R4"
/translation="MPEFSTINELMGSMVMIYLLLAGAGCWFTWRTGFIQPRYTRQFS
RSLKGSUSPQGGTSLAALTSALRIGSGNLAGVALIAAGPGPAGVFMVWVAIIG
MATSFACSLAOLYKERDPTQFRGPANMARGIMRWGMVVFALFLVLYAGLIFNS
VQANVASRALHFAFNIPPLISGLAFACALLIIIRGIGVARLWMLPLLIALLWVAG
SVFTCLWHIEOMPVGVIASIVKAFQWQEAAGAAGYTLTOAITSGCFORHMSNEAGMG
STPNAAATSTYPPHPVAQGVQMGVFDPTIICTSASAMILLIAGNASHSSTEGIQ
LLQHWSLTGEKASFVALIVILFAPSSIVANIIYAENNLFFLRHNAKAIWLLRLA
TLGMVAGTLISFLIWLQADMIMACMAITNLITALLSPVYITLAGDYLKORLIGVR
Query Match      23.2%; Score 335.8; DB 1; Length 251050;
Best Local Similarity 56.2%; Pred. No. 1.5e-31;
Matches 750; Conservative 0; Mismatches 552; Indels 33; Gaps 5;

QY 69 GACCACTTCGGATCGTGTGGCCCTCGGCAAGACGGCGATGTCCTCGTGGCTACCTG 128
DB 148829 GACAAAACGTCGTCATTCGCTGGGCTTAACCGGACTCTCTTGGCTGGACTTTT 148888
QY 129 GCGCGCCGCGCTTTCGCGTGTGCGATACCGAGAGAACCCGCGGAGCTGGCC 188
DB 148889 CTCGCCCGCGGCTGACCGCGGGGTGATGNACTCCGCTGACCGCCCGCGGTCTGGAT 148948
QY 189 ACCCTGCGTCCCACTATCCGAGGTGGAAGTGCCTTGCAGCGAACTCGAGCCGAGTTC 248
DB 148949 AAGCTGCCG-----CAAGAGGTGAGCGTCACTGCGGCGCTGAAGACGAGTGG 148999
QY 249 CTCCTGCTCGCCGCGAATCTATGTCAGCCCGCGCTGTGCTGCGGACCCCTCGCGTG 308
DB 149000 CTCCTAGCGCGGAGTTAATCGTCGCCAGCCCTGTGATGTGCTGCGCGCATCCCTCTCT 149059
QY 309 GTACAGGCGCGCGGAAAGCGTGCATCTCCGCTGACATCGATCTCTTCCCGCGCGAG 368
DB 149060 AGCGTGCAGCGCGTGGAGTGAATTTGCGCGGATATCGAACTGTTTTCGCGCGAA 149119
QY 369 GCGAAGGCGCGGATCGTCCCATCACCGGTTCCAAACGGAAGACCGCTGACCAACCCCTG 428
DB 149120 GCGAAGCGCGGATGTTGGCCATCACCGGCTCGAAGCGCAAAAGCACCGTACCACTTA 149179
QY 429 GTGGCGGAATGGCGGTGGCGCGGCAAGCGTGTGCGCGTGGCGGCAACCTCGGCAAC 488
DB 149180 GTGGCGGAGATGGCGAAGCGCGCGCGCTCAATGTGCGCGTGGCGCAATATCGCGTCT 149239
QY 489 CCGGCGCTCGACTCTCTGCGCGACGACATCGAGCTGTACGTGTTGAGTGTGCGAGTTC 548
DB 149240 CCGGCGCTGTGCTGCTGATGCGGATCGCGAATGTACGTGCTGGAATATCCAGTTTC 149299
QY 549 CAGCTGGAAACCTCGCATCGCTCAACCGGAGTGGCGACCGTGTCTGAACGTCAAGCGAA 608
DB 149300 CAATTGGAAACTACTCAAGTTTGCAAGCGCGCGCAACCGTGTCTCAAGTCACTGAA 149359
QY 609 GACCATATGATCGCTA---CGACGGCATGCTGACTACCACTGGCGCAAGACCGGATC 665
DB 149360 GATCATATGACCGGTATCCCTTTGTTGCAACAGTACCGCGGCGCGCAAACTGCGCGTC 149419
QY 666 TTCGCGGTGCGCGCGAGTGTGCTGATCGCGCGGATGCCCTGACCCGCGCTGATC 725
DB 149420 TACGAGAGGCGAAAGTGTGCGTGTGATGATGCGGATGCGGATGCGGATGCGGATGCG 149479
QY 726 GCGCATACCGTGGCTGTGCTGCTTCCGCTGGAACAGCGGAGCTTCAAGGCTTTCGCG 785
DB 149480 GGGGCGGATGAGCGCTGCTGCTAGCTTTGCGCTCATATGGGTGATATCA-----C 149530
QY 786 CTGATCGAGGAGACGCGCAGAGTGGCTGCGGTTCCAGTTCGCAAGCTGCTCGCGGTT 845
DB 149531 CTTAATCTGACGAGCGGCAACCTCGCTACGAGTCAAAAGTGAAGAGTGTGTAACCTG 149590
QY 846 GCGCAACTGAAGATCCGTGGCGCCACCACTATTCCACGCGCTCGCGCGCTGCGCGTG 905
DB 149591 AAAGAGATGAAGCTTCCGCGTCAGCATATACTACCAATGCGTGTAGCGCGCTGCGCGTG 149650

```

```

QY 906 GGCATCGCGTGGCGCTCCGCTTTCGACGCCATCGCTCGCGCGCTGAAGCGGCTTTCCCGC 965
DB 149651 GCGGATCGCGTAGGTCTCGCGCGCGCCAGCAGTTTGAAGCGCTTTCGACGACATTTACCGGT 149710
QY 966 CTGGCTCATCGCTCGCATGGGTAGCGAGCGGCGAGGCGGTGAGTCTACTACGACGATTCC 1025
DB 149711 CTGGCGCATCGCTTCCAGCTGCGGTGGAGCATAACGCGGCTTCTGATCAACGACTCG 149770
QY 1026 AAGGCGACCAAGCTCGCGCGCGCTTTCGCGGCGATCGAGGGGCTGGTGCCGACATCGAC 1085
DB 149771 AAAGCGACCAATGTCGGCAGTACCGAGCGGCGCTTAACGGTTTGCATG-----TGGAC 149824
QY 1086 GCGAAGCTGGTCTCTCGCGCGCGAGACGCAAGGCGCGGATTTCCATGACCTCGCG 1145
DB 149825 GGTACGCTGCATCTCTGCTGGCGCGGACGAGTAAGTCGGCAGACTTTTCTCCGCTGGCG 149884
QY 1146 GAGCGCGTGGCGGCTTCTGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1205
DB 149885 CGCTATCTGACCGCGGATGCTATCGCCTGTATTGCTTTGGCGCGATGGCGCGAGCTT 149944
QY 1206 GCGGAGGCACTGGGCAACGCGGTACCGTGGTGGCGCTCGCAACGCTGGACGAACGAGTC 1265
DB 149945 GCGGCACTGC-----GTCCGGAATCGCCCAAGACTGAGACGATGGAAGAGCGGATG 149998
QY 1266 GCGGAGGCGCGGAGCTGGCGCGGAGCGGATGCGGTGCTGCTGCTGCTGCTGCTGCTGCTG 1325
DB 149999 GCTTGTGCTGGCGCGGCTTCAGCGGCTGATATGCTGCTGCTGCTGCTGCTGCTGCTG 150058
QY 1326 AGCCTGGACATGTTCAAGAACTTCGAAGAACTCGAAGAACGCGGCGCTTTCGCCAAACCGTAGAG 1385
DB 150059 AGCCTGCATGATGTTTAAATAATTTGAGCAACGCGGCGGATGCTTTACCGCTGTGCGGAAG 150118
QY 1386 GAGCTAGCGTGTATGC 1400
DB 150119 GAGTGTAGTGTATGC 150133

RESULT 6
AE016834
LOCUS      300169 bp      DNA      linear      BCT 21-MAR-2003
DEFINITION      Salmonella enterica subsp. enterica serovar Typhi Ty2, section 1 of
ACCESSION      AE016834 AE014613
VERSION        AE016834.1 GI:29136167
KEYWORDS
SOURCE        Salmonella enterica subsp. enterica serovar Typhi Ty2
ORGANISM      Salmonella enterica subsp. enterica serovar Typhi Ty2
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Salmonella.
REFERENCE      1 (bases 1 to 300169)
AUTHORS        Deng, W., Liou, S. R., Plunkett III, G., Mayhew, G. F., Rose, D. J.,
Burland, V., Kodoyianni, V., Schwartz, D. C. and Blattner, F. R.
TITLE          Comparative Genomics of Salmonella enterica Serovar Typhi Strains
JOURNAL        Ty2 and Ctr18
MEDLINE        J. Bacteriol. 185 (7), 2330-2337 (2003)
PUBMED        22531367
                12644504
AUTHORS        2 (bases 1 to 300169)
                Deng, W., Liou, S. R., Plunkett, G. III, Mayhew, G. F., Rose, D. J.,
                Burland, V., Kodoyianni, V., Schwartz, D. C. and Blattner, F. R.
TITLE          Direct Submission
JOURNAL        Submitted (25-SEP-2002) Laboratory of Genetics, University of
                Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA
FEATURES
source        1. .300169
                /organism="Salmonella enterica subsp. enterica serovar
                Typhi Ty2"
                /mol_type="genomic DNA"
                /strain="Ty2"
                /serovar="Typhi"
                /sub_species="enterica"
                /db_xref="taxon:209261"
                190. .255

```

```
/gene="thrL"
/locus_tag="t0001"
190..255
CDS
/gene="thrL"
/locus_tag="t0001"
/notes="corresponds to STY0001 from Accession AL513382:
Salmonella typhi Ctl8"
/codon_start=1
/transl_table=11
/product="thr operon leader peptide"
/protein_id="AA067735.1"
/db_xref="GI:29136168"
/translation="MNRISTTTTITTTGNGAG"
337..2799
gene
/gene="thrA"
/locus_tag="t0002"
337..2799
CDS
/gene="thrA"
/locus_tag="t0002"
/notes="corresponds to STY0002 from Accession AL513382:
Salmonella typhi Ctl8; homoserine dehydrogenase I"
/codon_start=1
/transl_table=11
/product="aspartokinase I"
/protein_id="AA067736.1"
/db_xref="GI:29136169"
/translation="MRVLKFGGTSVANAERFLRVADILESNSRQGOVATVLSAPAKIT
HLVAMIEKTIGGDALPNISDAERIFSDLLAGLASAOPGFLPLARLKMVVEQFAQIK
HVLHGSLIGQCPDSINAAALICGEMKSTIAGLLEARGHRTVYIDPVEKLAVGHI
LESTVDIASTRIRASQIPADHMLMAGFTAGNEGELVILGRNGSDISAAVLAACL
RADCEIWDVDCDPRQVPDARLLKMSQYEAELSYGAKVHLHRTTPTIAOF
QPCLIKNTGNQAPLTIGASDDNLPVKGISNINMMFSGPGMKMGITGAAR
VFAAMRAGISVVLITQSSSEYSISFCQSCARARARADQDEFYLEKGLLEPLAV
TERLAISVVGDCMRTLRGISAKFFAALARANINIVAIAQGSSEISISVVVNDATT
GVRVTHOMLVDOIEVEVIGVGVGGLLEQLKQQTWLNKXHDILVCGVANSKA
LNTVHGLNDNQALAQANAPFNGLRILRLVKEHLNPIVIVCTSSQAVDQIAD
FLREGHVTPNKNANTSSMDYTHQURFAAASRRKFLYDNTNVAGLPLVNIENLLN
AGDELQKFGSLGSLGSLSFIFGKLEEGMSLQATALAREMGYTEPDPDDLSGMDVARK
LTLARETRELSDIVIEPVLPEFDSAGDVTFAPMAHLQDLDFAAARAKARDEG
KDVYVNTIEDEGVCVKIAEVDGNDPLFKVKNENALAFYSHYQPLPLVLRGYGAG
NRYTAGVFPADLLRLTSWLKV"
2801..3730
gene
/gene="thrB"
/locus_tag="t0003"
2801..3730
CDS
/gene="thrB"
/locus_tag="t0003"
/notes="corresponds to STY0003 from Accession AL513382:
Salmonella typhi Ctl8"
/codon_start=1
/transl_table=11
/product="homoserine kinase"
/protein_id="AA067737.1"
/db_xref="GI:29136170"
/translation="WKVYAPASSANMSVGFVLGAATVPDGTLLGDVVSVAEADHF
RLNHLGRFADKLPEPRENIVYQCWERFCQALGKTIIPVAMTLEKNPIGSGLGSSACS
VVAALVAMNEHCGLPNDTRLALMGELEGRISGSHYDNVAPCFGLGMLMEENGI
ISQVPGFDEWLVLPYKIVKSTAEARAILPAQYRQDCIAHGRHLAGFIHACYSGR
POLAAALMDVIAEPYRARKLLPGFSQARQAVSEIGALSGISSGGPTFLALCDKPETA
QRVADWLKXHLQNQSGEFVHCRLDTAGARVVG"
3734..5020
gene
/gene="thrC"
/locus_tag="t0004"
3734..5020
CDS
/gene="thrC"
/locus_tag="t0004"
/notes="corresponds to STY0004 from Accession AL513382:
Salmonella typhi Ctl8"
/codon_start=1
/transl_table=11
/product="threonine synthase"
/protein_id="AA067738.1"
/db_xref="GI:29136171"
/translation="MKLYNLKDHNEQVSAQAQVTOGLGKQOGLFFPHDLPEFSLTEID
EMLNQDFVSRSAKILSAFISGDETPQOILREVRPAAPAPVAQVSDVGCLELFRGP
TLAFKDFGGRFMAQMLTHISGDKPVTILTSGDTGAANAHAAYGLNVRVILYPRG
KISPLEKFLCTLGGNIEVAIDGDEDAQALYKQAFDDEELATGLNSANSINISR
LIAQCIYFEAQAQLPQAGARNQLVISVPSGNFGDLTAGLAKSLGSLPVKRFIAATNIN
DTPVRFHDGKAPKATQATLSNAMDSQPNMNPVDELFRRKIWRILTELGYYAAVDOT
TTQOTMRELKAGYISEPHAAYVRAIRDQLNPGYGLFLGTAPAKFKESVESILGE
TLALPEALAEADLPILLSHLHPADFAALRKLMTRQ"
complement(5114..5887)
-gene
/gene="yaaA"
/locus_tag="t0005"
complement(5114..5887)
CDS
/gene="yaaA"
/locus_tag="t0005"
/notes="corresponds to STY0005 from Accession AL513382:
Salmonella typhi Ctl8"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AA067739.1"
/db_xref="GI:29136172"
/translation="MLILISPAKTLDYQSLATTTRYTQPELDDHSHOOLIQARQLSAP
QISRLMGISDKLADLNATRFHDMQPHFTPDNAQALAFKGDVITGLQETFDNADF
FAQQHLRMLSGLYLRLDLMQPYRLENGIRLENPRKDLQYQFGDIITDKLINEALE
AQQDRVYNVLAEEYFKSVKPKKDLAELIKPVFLDEKNGKFKVYSYAKKARGLMGRF
IENRLTKPEQLTAFDREGYFFDEETSTQDELVFKRYEQ"
complement(5966..7396)
gene
/gene="yaaJ"
/locus_tag="t0006"
complement(5966..7396)
CDS
/gene="yaaJ"
/locus_tag="t0006"
/notes="corresponds to STY0006 from Accession AL513382:
Salmonella typhi Ctl8"
/codon_start=1
/transl_table=11
/product="putative amino-acid transport protein"
/protein_id="AA067740.1"
/db_xref="GI:29136173"
/translation="MPEFFSFNIEILWGSVMYLLLAGCWFRTWRTGTFQFRVIROFS
RSIKGSLSPQGLTSTFQALCTSLAARGSNLGVAGALATAGGCGAVFMWVSALIG
MATSFAECSLAQLYKERDPTGPRGPATMARGLNRMWGVFALFLVAGLIENS
VQANAVSRALHFAPNIPPLISGLAFAFCALLIIRIGKVGARLMQWLILIALLVAG
SVFICLWHIEQMPGVIAISVKSFAQMOEAAGAAGYTLTQAITSGFORGMFSNEAGM
STPNAATAATSYPPHPVAQIVQMGVSDTIIICVASAMILLAGNHASHSHTEGIQ
LIQHAMVSLTGEWGSFVALIVLFAFSIVANYIYAENNLFFLRCHNKAIAWLLRLA
LTGMVITAGTLLISFPLIWLQADMTMACMAITNLTAILLSPVYVTLAGDYLQRKLGVR
PQDPDRFDIEPQLAPDTWDAASRD"
7665..8618
gene
/gene="talB"
/locus_tag="t0007"
7665..8618
CDS
/gene="talB"
/locus_tag="t0007"
/notes="corresponds to STY0007 from Accession AL513382:
Salmonella typhi Ctl8"
/codon_start=1
/transl_table=11
/product="transaldolase B"
/protein_id="AA067741.1"
/db_xref="GI:29136174"
/translation="MTDKLTSRQFTTVVADTGDIAAMKLYQPDATTNPSSLTNAAQ
IPEYRKLIIDAVAWAKQSSDRAQVVDATDKLAVNIGLEIKLVPGRISTEVDARLS
YDTEASIAKAKRIKLYNDAGISNDRILIKLASTWQIGRAAEQLEKINGCNLTLLFS
FAQARACAGVYLISSPFVGRILDWKYKANTDKKDYAPADPGVGVSVTEIYEVYKHGY
ETVWAGASFRNVGEILELAGCDRLTIAPALLKELAESEGAIEKRLKLSFSEVKARPERI
TBAEFQWQHQPMAVDKLDGIRKFAVDQEKLEKMGIDL"
8729..9319
gene
/gene="mog"
/locus_tag="t0008"
8729..9319
CDS
/gene="mog"
/locus_tag="t0008"
```

/note="corresponds to STY0008 from Accession AL513382:  
Salmonella typhi CT18"  
/codon\_start=1  
/transl\_table=11  
/product="molybdopter in biosynthesis Mog protein"  
/protein\_id="AA067742.1"  
/db\_xref="GI:29136175"  
/translation="MNTLRIGLVSISDRSSGVYQDKIPALBEWLASALTTPFELQ  
RLIPDEQALIEQTLCELVDEMSCHLVLTGTGTPARDVTPDATLAIDREMPFGBO  
MROISLRFVPTAILSRQGVIRKOALILNLPQPKSIKETLEGKADDSVSPGIFA  
SVYCIQLLDGPVVEPAEVAAPKPSARRENSS"  
complement(9376..9942)  
/gene="yaaH"  
/locus\_tag="t0009"  
complement(9376..9942)  
/gene="yaaH"  
/locus\_tag="t0009"  
/note="corresponds to STY0009 from Accession AL513382:  
Salmonella typhi CT18"  
/codon\_start=1

Query Match 23.28; Score 335.8; DB 1; Length 300169;  
Best Local Similarity 56.28; Pred. No. 1.5e-31;  
Matches 750; Conservative 0; Mismatches 552; Indels 33; Gaps 5;  
  
QY 69 GACCACTTCGGCATCGTTGTCGGCTCGGCAAGAGCGGCATGTCCCTGGTGGCTACCTG 128  
DB 148821 GACAAAACGTCGTCATTATCGGTCTGGCTTAACCGGCATCTCTTCGGTGGACTTTT 148880  
  
QY 129 GCGCGCGCGGCTTGCCTTTGCGCGGTGTCGATACCCGAGAGAACCGCGGAGCTGGCC 188  
DB 148881 CTCGCCGCGCGGTGAGCGCGCGGTGATGATGATCTACGCGTGAACCGCGCGGTCTGGAT 148940  
  
QY 189 ACCCTGCGTCCAGTATCCGACGTGAAGTGTGCGCGCAACCTCGACGCGGAGTTC 248  
DB 148941 AAGTCGCG -----CAGAGGTGAGCGTCACGTTGGCGCCTGGAACGAGAGTGG 148991  
  
QY 249 CTCCTGCTCCGCGCGCACTATGTCAGCCCGCGCTTGTGCTGCGCACCCCTCGCGCTG 308  
DB 148992 CTCCTAGCGCGGATTAATCGTCGCGCAGCCCTGTTATGCTCTGCGCATCCCTCTCT 149051  
  
QY 309 GTACAGCGCGCGGAAAGCGTCCGATCTCCGGTGCATGTCATCTTCGCGCGCGGAG 368  
DB 149052 AGCGCTCCGCGCAGCGTGGAGTGAATTTGTCGGCGATATCGAATGTTTTCGCGGAA 149111  
  
QY 369 GCGAAGCGCGGATCGTCGCGCATCACCGGTTCCAAACGGAAGACCGGTGACACACCTG 428  
DB 149112 GCGCAAGCGCGGATGTGGCGATCACCGGCTCGAAGCGGAACACCGGTGACACACCTA 149171  
  
QY 429 GTGGCGGAAATGGCGGTGCGCGGACAAAGCGTGTGCGCGTGGCGGCAACCTCGGCACC 488  
DB 149172 GTGGCGGAGATGGCGGAGCGCGGCGTCAATGTCGCGTGGCGGCAATATCGGTCT 149231  
  
QY 489 CCGCGCTCGACCTGCTGGCGGACGACATGAGCTGTACGTGTGGAGCTGTGCGAGTTC 548  
DB 149232 CCGCGCTGTGATGCTGCTGGATGCGGATGCGGAAATTTGACGTGTGGAATATTCAGTTT 149291  
  
QY 549 CAGCTGGAACCTCGCATCGCTCAACCGCGAGTGGCGCGCTGGAAGTGTGAGCTCAGCGAA 608  
DB 149292 CAATTGGAACTACTCAAGTTTCAAGCGCGCGCGGACGAGTGTCAAGCTGACCTGAA 149351  
  
QY 609 GACCATATGATCGCTA---CGACGGCATGGCTGACTACCACTGGCCCAAGCACCGGAT 665  
DB 149352 GATCATATGACCGGTATCGTTGGTTTGAACAGTACCGGCGCGGCAACTCGCGGTC 149411  
  
QY 666 TTCGCGGTGCGCGCGAGTGTGGTGAATCGCGCGCATGCCCTGACCGCGCTGATC 725  
DB 149412 TACGAAGCGGAAAGTGTGCTGTAGTGAATGCCGATGACGCGTGTGACGATGCGGTACG 149471  
  
QY 726 GCCGATACCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 785  
DB 149472 GGGCGCATGAGCGCTCGGTGAGTGTGGTGTCAATATGGGTGATATCA-----C 149522

QY 786 CTGATCAGGAAGACGGCCAGAAAGTGGCTGGCTTCCAGTTCGACAAAGCTCTCGCGGTT 845  
DB 149523 CTTAATCCTCAGCAGCGGAAACCTGCTACGAGTCAAAGGTGAGAAGTCTCTGAACGTG 149582  
  
QY 846 GCGCAACTGAAGATCCGTGGCGGCCCAACAATATTTCCAAACGGCTCGCGCGCTGCGCGTG 905  
DB 149583 AAAGAGATGAAGCTTTCCGCTCAGCATTAACATATATACGCTTAGCGCGCTCGCGCTG 149642  
  
QY 906 GCGCATCGGTGCGCTTCCGCTTCCAGCCATGTGCGCGCGCTGAAGCGCTTTCCGCG 965  
DB 149643 GCGGATCGGTAGTGTGCGCGCGCGCAGCATGTTGAAGCGTTGACGACATTTACCGT 149702  
  
QY 966 CTGCTCATCGCTCCAGTGGGTACGACGAGCGGCGCTGAGCTACTACGACGATTC 1025  
DB 149703 CTGCGCATCGCTTCCAGCTGGCGTTGGAGCATTAACGCGCTTGGATCAACGACTCG 149762  
  
QY 1026 AAGCCACCAACGTCGGCGCGCGCTGCGCGCGATTCGAGGCGCTGGGTGCGGACATCGAC 1085  
DB 149763 AAACCGACCAATGTGCGCAGTACCGAAGCGCGCTAAACGGTTTGCATG-----TGGAC 149816  
  
QY 1086 GGAAGCTGTGCTGCTCGCGCGGAGAGCGGCGCGGCTTCCATGACTGCTGCGC 1145  
DB 149817 GGTAGCTGATCTGCTGCTGGCGGCGGACGCTAGTTCGCGGAGACTTTTCTCCGCTGCGC 149876  
  
QY 1146 GAGCGGTGCGCGCTTCTGCGCGCGGTGTACTGCTTGGCGCTGAGCGCGGCTGATT 1205  
DB 149877 CGCTATCTGACCGCGCATCGTATCGCGCTGATTGCTTTGGCGCGATGCGCGCAGCTT 149936  
  
QY 1206 GCCCAGGACTGGCAACGCGGTACCGCTGTGCGGTGCGTCAACGCTGAGCGGAGCAGTTC 1265  
DB 149937 GCCCAGCTGC-----GTCCGGAATCGCCCAACAGACTGAGACGATGAAGAGCGGATG 149990  
  
QY 1266 CCGCAGCGCGCGAGCTGGCGCGGAGCGCATCGGTGCTGCTGCTGCGCGCTGCGCG 1325  
DB 149991 CGTTGCTGGCGCGCAGCTTCAGCGCGGTGATGTCGTCGTCGTCGTCGTCGTCGTCG 150050  
  
QY 1326 AGCTGACATGTTCAAGAACTTCAAGAACCGGACGCGCTTGTTCGCCAAAGCCGTAGAG 1385  
DB 150051 AGCCTCATGATTTAAATAATTTGAGCAACGGCGCATGCTTTACCCGCTGCGCGAAG 150110  
  
QY 1386 GAGTACGGTGATGC 1400  
DB 150111 GAGTACGGTGATGC 150125  
  
RESULT 7  
AE008699  
LOCUS Salmonella typhimurium LT2, section 7 of 220 of the complete genome. BCT 23-APR-2003  
DEFINITION Salmonella typhimurium LT2, section 7 of 220 of the complete genome.  
ACCESSION AE008699 AE006468  
VERSION AE008699.1 GI:16418608  
KEYWORDS  
SOURCE Salmonella typhimurium LT2  
ORGANISM Salmonella typhimurium LT2  
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Salmonella.  
REFERENCE 1 (bases 1 to 22348)  
AUTHORS McClelland, M., Sanderson, K.E., Spieth, J., Clifton, S.W., Hou, S., Layman, D., Leonard, S., Ali, J., Dante, M., Du, F., Grewal, N., Mulvaney, E., Ryan, E., Sun, H., Florea, L., Miller, W., Stoneking, T., Nhan, M., Waterston, R. and Wilson, R.K.  
Complete genome sequence of Salmonella enterica serovar Typhimurium LT2  
TITLE  
JOURNAL Nature 413 (6858), 852-856 (2001)  
MEDLINE 21534948  
PUBMED 11677609  
REFERENCE 2 (bases 1 to 22348)  
AUTHORS  
CONSRPM  
TITLE The Salmonella typhimurium Genome Sequencing Project Direct Submission  
JOURNAL Submitted (29-MAR-2001) Genome Sequencing Center, Department of

Genetics, Washington University School of Medicine, 4444 Forest Park Boulevard, St. Louis, MO 63108, USA  
 COMMENT Supported by NIH grant 5U 01 AI43283

## COMMENT

Coding sequences below are predicted from manually evaluated computer analysis, using similarity information and the programs: GLIMMER; <http://www.tigr.org/softlab/glimmer/glimmer.html> and GeneMark; <http://opal.biology.gatech.edu/GeneMark/>

EC numbers were kindly provided by Junko Yabuzaki and the Kyoto Encyclopedia of Genes and Genomes; <http://www.genome.ad.jp/kegg/>, and Pedro Romero and Peter Karp at Ecocyc; <http://ecocyc.PangeaSyste.ms.com/ecocyc/>

The analyses of ribosome binding sites and promoter binding sites were kindly provided by Heladia Salgado, Julio Collado-Vides and ReguonDB; [http://kinich.cifn.unam.mx:8850/db/regulondb\\_intro.frameset](http://kinich.cifn.unam.mx:8850/db/regulondb_intro.frameset)

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistries or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one m13 subclone.

## FEATURES

source  
 1. .22348  
 /organism="Salmonella typhimurium LT2"  
 /mol\_type="genomic DNA"  
 /strain="LT2; SGSC 1412; ATCC 700720"  
 /db\_xref="ATCC:700720"  
 /db\_xref="taxon:99287"  
 /note="LT2"  
 /gene="leuB"  
 /note="synonym: STM0110"  
 complement(298. .915)  
 /EC\_number="4.2.1.33"  
 /note="3-isopropylmalate dehydratase small subunit."  
 (SW:LEU2\_SALTY)  
 /codon\_start=1  
 /transl\_table=11  
 /product="3-isopropylmalate isomerase (dehydratase), subunit with LeuB"  
 /protein\_id="AAL19074.1"  
 /db\_xref="GI:16418609"  
 /translation="NAEKFTQHTGLVVPDAAVDTDAIIPQFLQKVTTCFGAHLF  
 NDWRFLDEKQGNPEFVNFPEYGCASILLARENFGCGSSREHAFWALTDYGRKVI  
 AFSFADFTGNSFNQLLPVTLSDAQVDFLALVKANPGIKFEVDFLEAQVRAKDKTY  
 SPKIDDFRRHCLMGIDSLGLTLQHEDAIAAYENKQAPFMR"  
 complement(910. .915)  
 /gene="leuB"  
 /note="putative RBS for leuB; RegulonDB:STMSIH000490"  
 complement(914. .2327)  
 /gene="leuB"  
 /note="synonym: STM0111"  
 complement(914. .2314)  
 /gene="leuB"  
 /EC\_number="4.2.1.33"  
 /note="3-isopropylmalate dehydratase large subunit."  
 (SW:LEU2\_SALTY)  
 /codon\_start=1  
 /transl\_table=11  
 /product="3-isopropylmalate isomerase (dehydratase), subunit with LeuB"  
 /protein\_id="AAL19075.1"  
 /db\_xref="GI:16418610"  
 /translation="NAKLYEKLFDHVVFEAPNETPLLYIDRHLVHEVTSPOAFDGL  
 RAHRPQPGKTFATMDHNVTQTKDINASGEMIOELIKNCEFGELYDLNLH  
 PYQGLVHYMGPPQVTLPCMTIVCGDSHTAHGAFALAGLTSEVHYLATQTLKQ  
 GRATKMLEVGNAPGITAKDIVLAIGKYGSGAGTGHHVEFCGDALRALSMEGRMT  
 LCNMAIEMGAKAGLVAPDETTFNFKVGRHAPKGRDFDEAVEYKNTLTKDDGATEDTY

VALRAEEIAPQVTWGTNPQOIVISVTDIIPDPASFSDPVERASAEEKALAYMQLQPGVPL  
 TDVAIDKVFISGTSNRIEDRAAAAEVAKRKVAPGVQALVYPGSPVKAQAEAGLD  
 KIFIEAGFEWRLLPGGSMCLAMNDRLNPGERCASTNRNFEGRQGRGGRTHLVSPAMA  
 AAAAVTGHFADIRSIK"  
 complement(2317. .3422)  
 /gene="leuB"  
 /note="synonym: STM0112"  
 complement(2317. .3408)  
 /gene="leuB"  
 /EC\_number="1.1.1.85"  
 /note="3-isopropylmalate dehydrogenase. (SW:LEU3\_SALTY)"  
 /codon\_start=1  
 /transl\_table=11  
 /product="3-isopropylmalate dehydrogenase"  
 /protein\_id="AAL19076.1"  
 /db\_xref="GI:16418611"  
 /translation="MSKNYHIAVLPGDGGICPEVMAQALKAVMDAVSRFRDMRTTSHYD  
 VGGIADNHHPLPRATVEGCOADAILFGSGPKRWENLPPESEPERGALLPLRRHF  
 KLFNLNRPALYQGLEAFCLRADIAAAGFDILCVRELTTGGIYFGPKRGESGGYKE  
 ADTEYVHFREIERIARIAFESARKRRRTVITDKANVLQSSILNREIYNDVAKTVPD  
 VELAHMYIDNATWQIKDPSQDFDLSNLFGLDILSDCAMITGSMGLPSPASLNQOG  
 FGLYEPGAGSADPIAGKIANPFAQILSLALLRLYSLDANDAATAIEQAIRALESGV  
 RTGDLARGAAVSTDEMGDIARYVAEGV"  
 complement(2322. .2327)  
 /gene="leuB"  
 /note="putative RBS for leuB; RegulonDB:STMSIH000491"  
 complement(3408. .4999)  
 /gene="leuA"  
 /note="synonym: STM0113"  
 complement(3408. .4979)  
 /gene="leuA"  
 /EC\_number="4.1.3.12"  
 /note="similar to E. coli 2-isopropylmalate synthase  
 (AAC73185.1); Blastp hit to AAC73185.1 (523 aa), 92%  
 identity in aa 1 - 523"  
 /codon\_start=1  
 /transl\_table=11  
 /product="2-isopropylmalate synthase"  
 /protein\_id="AAL19077.1"  
 /db\_xref="GI:16418612"  
 /translation="MSQQVIFDITLDRGEQALQASLAKELQIALALERMGVDMVE  
 VGFVSSPGDFESVQTIARTIKNSRVCALRCVEKDIDVAQALKAVADAFRIHTFIAT  
 SPWHIATKLRLDVEIERAVYVWRKARYNTDDVEFSCDAGRTPVDDDLARVVEAAIN  
 AGARTINIPDTVTGTPPEFAGIISGLSERVNPNDKALISVHTDDGLIAGVNSLAAY  
 HAGAPVEGAMNGIGERAGNCALEEYVIAIKYKDKIMVNTNHNHIEIWRITSQVSI  
 CNMPTANKAIVGSAFAGSHGQDVLKKNREYIEMPTSEIGLNIQILNTSRSGR  
 AAVKRMEMGYKDTDYNDMLYDAFLKADKGOVDYDLALAFINKQOQEPHFPR  
 LDYFSVQSSSDIATASVKLAGEIKAEANGPVDAYIQAIRITCYDQVELVKYD  
 LNAKQKDALGQVDIVVNNHGRFHGVLATDIVESSAKAMVHVLNNIWRAREVEKE  
 LQKKAQNKENKQTV"  
 complement(3417. .3422)  
 /gene="leuB"  
 /note="putative RBS for leuB; RegulonDB:STMSIH000492"  
 complement(4994. .4999)  
 /gene="leuA"  
 /note="putative RBS for leuA; RegulonDB:STMSIH000493"  
 complement(5067. .5298)  
 /gene="leuB"  
 /note="synonym: STM0114"  
 complement(5067. .5153)  
 /gene="leuB"  
 /note="leu operon leader peptide. (SW:LPL\_SALTY)"  
 /codon\_start=1  
 /transl\_table=11  
 /product="leu operon leader peptide"  
 /protein\_id="AAL19078.1"  
 /db\_xref="GI:16418613"  
 /translation="MSHIVRETGLLLNAFIVRGRPVGGIQH"  
 complement(5162. .5167)  
 /gene="leuB"  
 /note="putative RBS for leuB; RegulonDB:STMSIH000494"  
 complement(5271. .5279)  
 /gene="leuB"

-10\_signal

[illegible]

K12 and O157  
Nucleic Acids Res. 30 (20), 4432-4441 (2002)  
12384590  
2 (bases 1 to 11441)  
Jin, O., Shen, Y., Wang, J. H., Liu, H., Yang, J., Yang, F., Zhang, X. B.,  
Zhang, J. Y., Yang, G. W., Wu, H. T., Dong, J., Sun, L. L., Xue, Y.,  
Zhao, A. L., Gao, Y. S., Zhu, J. P., Chen, S. X., Yao, Z. J., Wang, Y.,  
Lu, W. C., Qiang, B. Q., Wen, Y. M., and Hou, Y. D.  
Direct Submission  
Submitted (08-MAY-2001) Microbial Genome Center of Chinese Ministry  
of Public Health, 100 Yingxin Jie, Xuanwu Qu, Beijing 100052, P. R.  
China

FEATURES  
source  
Location/Qualifiers  
1. .11441  
/organism="Shigella flexneri 2a str. 301"  
/mol\_type="genomic DNA"  
/strain="301"  
/serotype="2a"  
/db\_xref="taxon:198214"  
292. .786  
/gene="yabB"  
/locus\_tag="SF0078"  
292. .786  
/gene="yabB"  
/locus\_tag="SF0078"  
/note="Residues 13 to 164 of 164 are 99 pct identical to  
residues 1 to 152 of a 152 aa protein from Escherichia  
coli O157:H7 EDL933 ref: NP\_285777.1"  
/codon\_start=1  
/transl\_table=11  
/product="orf, conserved hypothetical protein"  
/protein\_id="AAN41743.1"  
/db\_xref="GI:24050284"  
/translation="MGKVGIRGEAGMFGATVNLDSKRLSVPTRYREOLLENAAG  
QWCTTIDHHPCLLYLPEWEIIEQLKSLLSMNPYRVRQLLGHASECOMDAG  
RLIAPVLRHAGLTREYMLVQFNKFLWDETTHQVQKEDIDAEQLATGDSERLQ  
DLSL"  
788. .1729  
/gene="yabC"  
/locus\_tag="SF0079"  
788. .1729  
/gene="yabC"  
/locus\_tag="SF0079"  
/note="Residues 1 to 313 of 313 are 99 pct identical to  
residues 34 to 346 of a 346 aa protein from Escherichia  
coli pir: QBCFt"  
/codon\_start=1  
/transl\_table=11  
/product="yabc protein"  
/protein\_id="AAN41744.1"  
/db\_xref="GI:24050285"  
/translation="MNEYKHTTVLLDEAVNGLINRDPDGIYIDGTFGRGHSRLILSQ  
LGEGRLAIDRDPQAIYAKVTTDDPFSIIGHFSAIGEYVAERDLKIDGILLDL  
GVSPOLDAERGFSPRQGLDMRMDPTRGQSAEWLQTAEADIANVLKYGEHFP  
AKRIARIVNREQPMRTKELAEVVAATPVKRFKHPATRTFOAVRIWNSELEE  
IEQALKSSVNLAPAGRLSLTISPHSLIEDRIVKRFMRNSRGPPQVAGLPMTPEQLKL  
GGQLRALGKLMPEEVAENPRASVLRFAERTNA"  
1726. .2091  
/gene="ftsL"  
/locus\_tag="SF0080"  
1726. .2091  
/gene="ftsL"  
/locus\_tag="SF0080"  
/note="Residues 1 to 121 of 121 are 100 pct identical to  
residues 1 to 121 of a 121 aa protein from Escherichia  
coli O157:H7 EDL933 ref: NP\_285779.1"  
/codon\_start=1  
/transl\_table=11  
/product="cell division protein; ingrowth of wall at  
septum"  
/protein\_id="AAN41745.1"  
/db\_xref="GI:24050286"  
/translation="MISRVTALSKVKGSMGSHERHALPGVIGDILLRFGKLPCLFI

gene  
/gene="ftsI"  
/locus\_tag="SF0081"  
2107. .3873  
/gene="ftsI"  
/locus\_tag="SF0081"  
/note="Residues 1 to 588 of 588 are 99 pct identical to  
residues 1 to 588 of a 588 aa protein from Escherichia  
coli O157:H7 EDL933 ref: NP\_285780.1"  
/codon\_start=1  
/transl\_table=11  
/product="septum formation; penicillin-binding protein 3;  
peptidoglycan synthetase"  
/protein\_id="AAN41746.1"  
/db\_xref="GI:24050287"  
/translation="MKAATQKPKQEEHANFISWRFALLCCGCIILALAFILGRVAM  
LQVISPDLVKEGDMRSLRVQVSTSGRMITDRSGPLAVSPVPAIWDAPKREHDAG  
GISVGDWALANALNPLDQLSARINANPKGRFIYLAQVNSDMADYTKKLRGTH  
LREESRYYPGSEVTAHLIGFTNVDSQGTGVEKSPDKWLTGQGERIVRKDRYGRVI  
EDISSTDSQAANHLALSIDERLQALYRELNNAVAKAESGSAVLVDNTGVELAMA  
NSPSYNNLSGTPKEMRNRTIIDVEPGSTVYKPMVMTALQGRGVVRENSVUNTPIY  
RINGHEIKDVARSELTLTGVQKSSNVGSVKALTPSSALVDITYSRFGKATNLG  
LVYRSGLYPQKQWSDIERATFSFGYGLMTPLQIARVATIGSYGIRPLSTITVD  
PPVGRVFPESIVRTVWHMESVLPVGGGKAAIKGYRIATIKGTAKKVGPDGRYI  
NKVIYTAGVAPASOPREALVVVINDPQAGKYYGGAVSAPVFGAINGGVLRTWNIEPD  
ALTGDKNEVINQSGEGTGRS"  
3860. .5347  
/gene="murE"  
/locus\_tag="SF0082"  
3860. .5347  
/gene="murE"  
/locus\_tag="SF0082"  
/EC\_number="6.3.2.13"  
/note="Residues 1 to 495 of 495 are 99 pct identical to  
residues 1 to 495 of a 495 aa protein from Escherichia  
coli pir: S40595"  
/codon\_start=1  
/transl\_table=11  
/product="UDP-N-acetylmuramoylalanyl-D-glutamate--2,  
6-diaminopimelate ligase"  
/protein\_id="AAN41747.1"  
/db\_xref="GI:24050288"  
/translation="MADRNLRLLPWPDPAPSRLALREMTLDSRVAAGDLFVAVVGH  
QADGRYIPQAIQAQVAAITAEAKDEATDGEIREMHGVPIVYLSOLNERLSALAGRFY  
HPSDNRLVGYTGNGKTTTQLLAQWSQLGETSAVMGTGNGLKVIPTEHTTG  
SAVDQHELAGLVDDQATPCAMEVSHGLVQHRVAALKFPASVFTNLSRDHLDYHGDM  
EHYEAARWLLYSEHHCGQAIINADDEVRRRLAKLPDAVAVSMEDHINPNHGRWLKA  
TEVNYHDSGATIRFSSWGSGDEIETHLMGAFVSNLLALATLALGLYPLADLLKTA  
RLQPVGRMEVETAPGKTVTVVDYAHTPDALEKALQARLHCAGKLCVFCGGGDRDK  
GKPLMGTAEEFADVAVVTDNPRTEPEPRAIINDLAGMLDAGHAKVMEGRAEAVTC  
AVNQAKENDVVLVAGKGHEDYQIVGNQLDYSRVTVARLLGGIA"  
5344. .6702  
/gene="murF"  
/locus\_tag="SF0083"  
5344. .6702  
/gene="murF"  
/locus\_tag="SF0083"  
/note="Residues 1 to 452 of 452 are 99 pct identical to  
residues 1 to 452 of a 452 aa protein from Escherichia  
coli K12 ref: NP\_414628.1"  
/codon\_start=1  
/transl\_table=11  
/product="D-alanine:D-alanine-adding enzyme"  
/protein\_id="AAN41748.1"  
/db\_xref="GI:24050289"  
/translation="MISVTLSQLTDILNGLQAGADITLDAVTTDTTRKLLPFGCLFVALK  
GRFPAHDPAQAKAGGAGALLVSPDLIDLPLQIVKDRFLAFAGELAAVVRQVPAIV  
VALTSSGKTSVKQKMTAALISQCGNTLYTAGLNNDIGVPTLLRLLTPDYDVAVIELG  
ANHOGEIAWTVSLTRPEAALVNNLAAHLEGFGLAGVAKAGEIFSGLPENGIALMN  
ADNDLWLNQSVLTSRQKVRFPNARNSDFTATNTHVTSHGMEFTLQPTSGVSDLLP  
LPGRRNIANALAAALSMVSGATLDLDAIKAGLANLKNVGRLLPQIQAENQLLLDSSYN





```

REFERENCE
AUTHORS
1 (bases 1 to 290029)
Wei,J., Goldberg,M.B., Burland,V., Venkatesan,M.M., Deng,W.,
Fournier,G., Mayhew,G.F., Plunkett,G. III, Rose,D.J., Darling,A.,
Mau,B., Perna,N.T., Payne,S.M., Runyen-Janecky,L.J., Zhou,S.,
Schwartz,D.C. and Blattner,F.R.
Complete Genome Sequence and Comparative Genomics of Shigella
flexneri Serotype 2a Strain 2457T
Infect. Immun. 71 (5), 2775-2786 (2003)
12704152
JOURNAL
PUBMED
2 (bases 1 to 290029)
Wei,J., Goldberg,M.B., Burland,V., Venkatesan,M.M., Deng,W.,
Fournier,G., Mayhew,G.F., Plunkett,G. III, Rose,D.J., Darling,A.,
Mau,B., Perna,N.T., Payne,S.M., Runyen-Janecky,L.J., Zhou,S.,
Schwartz,D.C. and Blattner,F.R.
Direct Submission
Submitted (13-JUN-2002) Genetics Laboratory, University of
Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA
FEATURES
source
1. .290029
/organism="Shigella flexneri 2a str. 2457T"
/mol_type="genomic DNA"
/strain="2457T"
/serotype="2a"
/db_xref="taxon:198215"
190. .255
/genes="thrL"
/locus_tag="S0001"
190. .255
/genes="thrL"
/locus_tag="S0001"
/locus_tag="S0001"
/function="leader; Amino acid biosynthesis: Threonine"
/notes="residues 1 to 21 of 21 are 100 pct identical to
residues 1 to 21 of 21 from Escherichia coli K-12 : B0001"
/codon_start=1
/transl_table=11
/product="thr operon leader peptide"
/protein_id="AAP15548.1"
/db_xref="GI:30039814"
/translacion="MKRISTITTTITTTITGNGAG"
336. .2798
/genes="thrA"
/locus_tag="S0002"
336. .2798
/genes="thrA"
/locus_tag="S0002"
/function="enzyme; Amino acid biosynthesis: Threonine"
/notes="residues 1 to 820 of 820 are 98.04 pct identical to
residues 1 to 820 of 820 from Escherichia coli K-12 :
B0002"
/codon_start=1
/transl_table=11
/product="aspartokinase I, homoserine dehydrogenase I"
/protein_id="AAP15549.1"
/db_xref="GI:30039815"
/translacion="MRVLEGGTSVANERFLRVADILESNAQOGVATVLSAPAKIT
NHLVMIERTISGDALPNISDAERIFASLLTGLAAQPGFLAQKTFVDDQFAQIK
HVLHGSLGQCPDSINAAILICRGEKMSIAGVLEARGHNVTVIDPVKLLAVGHY
LESTDVIASTRIRASIRPADHVMFMAGFTAGNEKGELVLRNGSDYSAAVLAACL
RADCEIWTDDVGVTCDPQVPDARLLKMSYQEAELSFGAKVLHPRTTPTPAOF
QIPCLIKTGNQAPGTLIGASDEDELVPKGISLNNMAMFESVSGPMKGMVGMAR
VFAMRSARISVVLITQSSEYSISFCVPQSDCYRAERAMQEEFYLEKEGLEPLAV
TERLALISVGMRTLRGISAKFFAALARINIVAIAGSSERSISVVVNNDDATT
GVVTHOMLFNTDQVLEVEVIGVGGGALLQLKQKQKMLKNKHIDLRVCVANSKA
LITSVHGLNENWLELAQAKEPFNILRLVKEYFLNLPVIVCTSSQAVADQYAD
FLREGPHVTPNKNKATSMQYHOLRYAAEKSRKFLYDVTNAGLPIVENLQNLN
AGDELVKFSYILSGLSYIFGKLDGCMFSFSEATLTAREMGYTEPPRODLSGMDVAK
LILARETGRELEADIEIPEVLPREFNAGDVAFMANLSQLDDLFAARVAKARDEG
KVLRYVGNIDEGVCKIKAEVDGNDPLFKVNGENALAFYSHIYQIPLPLVRGYGAG
NDVTAAGVFADLLRTLSSWKLGV"
2800. .3732
/genes="thrB"
/locus_tag="S0003"
2800. .3732
/genes="thrB"
/locus_tag="S0003"
/locus_tag="S0003"
/function="enzyme; Amino acid biosynthesis: Threonine"
/notes="residues 1 to 310 of 310 are 95.48 pct identical to
residues 1 to 310 of 310 from Escherichia coli K-12 :
B0003"
/codon_start=1
/transl_table=11
/product="homoserine kinase"
/protein_id="AAP15550.1"
/db_xref="GI:30039816"
/translacion="MVKVYAPASANNMVGFDVLGAAPVGDALGILGDVVTVAAETFF
SNNLGRFADKLFPSEPRENVTCWRFQELGKQIPVAMTEKNMPICGSLSSACS
VVAALMAMNEHCGKPLNDRLIALMGELEGRISGSIHYDNVAPCGHGLQMLTEENDI
ISQOVPGFDEWLVLVAYPGIKVYSTAEARAILPAQYRQDQCIAGHRILAGFIHACYSRQ
PELAAKLMDVIAEPYRERILLPGFRQARQAVAEIGAVASGISGSGFTLALCDKPDPA
QRVADWLIGNYQNGEFGVHICKLDITAGARVLEN"
3733. .5019
/genes="thrC"
/locus_tag="S0004"
3733. .5019
/genes="thrC"
/locus_tag="S0004"
/function="enzyme; Amino acid biosynthesis: Threonine"
/notes="residues 1 to 428 of 428 are 95.56 pct identical to
residues 1 to 428 of 428 from Escherichia coli K-12 :
B0004"
/codon_start=1
/transl_table=11
/product="threonine synthase"
/protein_id="AAP15551.1"
/db_xref="GI:30039817"
/translacion="MKLYLKDHNHNEQVFAQVTOGLKNGQLFFPHDLPEFSLTEID
EMLKIDFVRSAKILSAFIDGDEIQEILREVRAPAFAPVAVNVESDVGCLLPHGP
TLAFDGGGFMAQMLTHIADDPVTILTATSDTGAAVAHAFFYGPVNVKVVILPRG
KISPGQKFLCTGLGNIETVAIDGDQACALYKQAFDDEELKVALGLNSANSINSR
LIAQICYFEVAQQLPQEARNLQVSVPSNGFDLTAGLLAKSFGLPVKRFIAATNVN
DTVPRLHDGQSPKATQATLSNAMDVQPNMNPVBEELFRKIMQLKELGFAAVDDE
TTQOTMRELKELGYTSEPHAAVAYRALDQLNPGEYGLGTGTAHPAKFESVEAILGE
TLDLPKELAEADLPLLSHNPADFALSRLKLMNHQ"
5242. .5529
/locus_tag="S0005"
5242. .5529
/locus_tag="S0005"
/notes="residues 1 to 65 of 95 are 60.00 pct identical to
residues 1 to 68 of 98 from Escherichia coli K-12 : B0005"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAP15552.1"
/db_xref="GI:30039818"
/translacion="MQSIVLALSVLVAPMAAAAEITLVPSVKLQIGDRDNRGVYWD
GGHWRDHGWWKQHYEWRGNRWHPGPPPPRHHKKAPHDHGGHGPCKHHR"
5564. .5668
/notes="REP (repetitive extragenic palindromic) element,
contains 2 REP sequences"
complement(5682..6458)
/genes="yaaA"
/locus_tag="S0006"
complement(5682..6458)
/genes="yaaA"
/locus_tag="S0006"
/notes="residues 1 to 258 of 258 are 100.00 pct identical
to residues 1 to 258 of 258 from Escherichia coli K-12 :
B0006"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAP15553.1"
/db_xref="GI:30039819"
/translacion="MLIISPAKTLDYOSPLTTRTYTLPELLDNSQOLIHEARKLTPP
OISTLMRISDKLAGINAAEFHDMQDFTPANARQAILAFKGDVYTGLOAETSEDDFD
FAQQLHRLMSGLYGLVRPLDLDMQPYRLENGIRLENARGKDLYQFWGDIITNKLNEALA

```





[illegible]

```

/function="enzyme; Cell envelop: Murein sacculus,
peptidoglycan"
/notes="Residues 1 to 495 are 98.98 pct identical to
residues 1 to 495 of 495 from Escherichia coli K-12 Strain
MG1655: B0085"
/codon_start=1
/transl_table=1
/product="meso-diaminopimelate-adding enzyme"
/protein_id="AAG54389.1"
/db_xref="GI:12512788"
/translacion="MADRLNLDLAPWDPAPSRALREMLDSRVAAGDLFVAVVGH
HEDSNLXQVIAQGTAAIAEAKGEATDGEIREMHGVPVYIYLSQNERLSALAGREY
SAVDQHELGLVQDQATFCAMEVSSHGLVQHRVAALKFAASVTNLSRDLHDYHGM
EYEAQKWLSEHHCQAIIINADDEVRRWLAKLPDAVAVSMEDHINPNCGRWLKA
TEVNHDSGATIRESSSGDGEIESHLMGAPVSNLLILALALCALGYPLADLLKTA
RLQPCVGMVFTAPGRPTVVYDYAHTPDALKALQARLHCAGKLCWFCGCGDRK
KLPMLGAIAEFDVAVVDDNPRTEPRRAIINDILAGMLDAGHAKVMEADRAEAVTC
AVMQAKENDVVLVAGKHEDYQIVGNQRLDYSRVTVARLLGVIA"
6994..8052
/gene="murF"
/notes="synonym: 20096"
6994..8052
/gene="murF"
/function="enzyme; Cell envelop: Murein sacculus,
peptidoglycan"
/notes="Residues 1 to 452 of 452 are 99.33 pct identical to
residues 1 to 452 of 452 from Escherichia coli K-12 Strain
MG1655: B0086"
/codon_start=1
/transl_table=1
/product="D-alanine:D-alanine-adding enzyme"
/protein_id="AAG54390.1"
/db_xref="GI:12512789"
/translacion="MISVTSQLDILNGLQAGADITLDAVTTDRKLTGCLFVALK
GERDADFADQAKAGAGALLVSRDIDPLIVKDRFLAFGELAARWQQVPARY
VALTSGSGSKEMTAAILSOCCNTLYTAGNNDIGVPMPTLLRLPEYDYAVIEG
ANHGEIVWTVSLTRPAAIIVNNLAAHLEGFGLAGVAKAGEIFSLGPENGIAIN
ADNDMLNWSVIGSRVWRFSNPAANSDFATNIHVTSHGTEFTLOTPTGSDVLLP
LPGRHNTANAALAAALSMVGATLDAIKAGLANIKAVPGRFPLOLAENOLLDDSYN
ANVSGMTAAVLEMPGKRVLVVGDMAELGAESEACHVQVGEAAKAGIDRVLSVGK
QSHAISTASGVGEHFAEKTALITRLKSLIAEQVITILVSGSRAAAEVVRALQENG
TC"
8046..9128
/gene="murY"
/notes="synonym: 20097"
8046..9128
/gene="murY"
/function="enzyme; Cell envelop: Murein sacculus,
peptidoglycan"
/notes="Residues 1 to 360 of 360 are 99.72 pct identical to
residues 1 to 360 of 360 from Escherichia coli K-12 Strain
residues 1 to 360 of 360 from Escherichia coli K-12 Strain
Query Match 19.9%; Score 288; DB 1; Length 12791;
Best Local Similarity 53.5%; Pred. No. 1.5e-25;
Matches 734; Conservative 0; Mismatches 605; Indels 33; Gaps 5;

Qy 82 TCGTTGTCGGCTCGGAAGAGCGCATGTCCTGGTGCCTACCTGGCGCGCCGCGCT 141
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 142 TGCCTTTGCGGTGTCATACCCGAGAAACCCCGGAGCTGGCCACCCTGGCTGCC 201
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 9219 TGACGCCGCGGTATGATACGCGTATGACACCGCTGGCTGGATAAATACCGG 9275
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 202 AGTATCCGAGTGAAGTCGTTGGCGGCAACTGACGCCGAGTTCCTCTGCTCGGCC 261
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 9276 -----AAGCCGCTAGACGCCACACGGGCACCTCGAATGATGAATGGCTGATGGCGGCG 9329
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 262 CGCAACTCTATGTAGCCCGCGCTGTGCTGCGCAGCCCTGCGCTGGTACAGCGCGCG 321
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 9330 ATCTGATTGTGCCAGTCCCGGTATTGCACTGGCGCATCCATCTTAAGCGCTGCCGCGTG 9389
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Qy 322 CGAAGGGCGTCCGATCTCCGGTGCATCGATCTCTTCGCCCGCGAGGCGAAGCCCGGA 381
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 9390 ATGCGGAATCGAAATCGTTGGCGATATCGAGCTTCTGTGCGGAGCAACACCGGA 9449
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 382 TCGTCGCCATCACCGTTTCCAAACGCGAAGAGACCGTGCACCCCTGGTGGCGGAATGG 441
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 9450 TTGTGGCATTACCGGTTTCTAACGCGAAAGACCGTGCACACGCTAGTGGGTGAATGG 9509
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 442 CGGTGGCGCGGACAAAGCGTGCAGCGTGGCGGCAACCTCGGCAACCCCGCGCTCGACC 501
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 9510 GAAAGCGCGGGGGTTAAACGTTGGTGGTGGCAATATTGGCCCTGCCCGGTTGATGC 9569
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 502 TCGTGGCGAGCATCGAGCTGTACGTGTGGAGCTGTGAGAGCTTCAGCTTCCAGCTGGAACCT 561
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 9570 TGTGGATGATGAGTGGCACTGTACGTACTGGAACCTGTGAGCTTCCAGCTTCCAGTGAACCA 9629
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 562 CGATTCGCTCAACGCCAGGTGGCGACCGTGTGAGCTGCAGCGTGCAGCAAGACATATGGATC 621
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 9630 CTTCCAGCTTACAGCGGTAGCAGCGACCATCTTGACGTGACTGAAGATCATATGATGATC 9689
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 622 GCTA--CGAGGCGATGCTGACTTACCACTGGCGCAACGCGGATCTTCCGCGGTGCC 678
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 9690 GCTATCCGTTTGGTTTACAGCAATACCGTGGCGCAAAATTTGGTATTACGAAAATGCGA 9749
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 679 GCCAGTCTGCTGAATCGCGCCGATGCCCTGACCGGACCGCTGATCCCGATACCGTGC 738
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 9750 AAGTTTGTGCTCAATGCTGATGATGCTTAAACATGCCGATTCGCGTGGGATGAAC 9809
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 739 CGTCTGCTGCTTGGCGCTGAACAAAGCCGAGCTTCAAGCGCTTTCGCGCTGATCGAGGAAG 798
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 9810 GCTGCGTCAAGCTTTGGCGTCAACATGGGTGACTATCA-----CCTGAATCATCAGC 9860
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 799 ACGGCGAAGTGGTGGCTTCCAGTTCGCAAGCTGCTGCCGCTTGGCGCAACTGAAGA 858
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 9861 AGGCGAAACCTGGCTGGGTGAAAGCGAGAAAGTCTGAACGTGAAAGAAATGAAC 9920
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 859 TCCGTGGCGCCACAACTATTCCAAACGCGCTCGCGCTGGCGCTGGCGCATCGGCTCG 918
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 9921 TTTCCGGGACAGATAACTACACCAATGCGTGGCGGCGCTGGCGCTGGCAGTCCGCGAG 9980
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 919 GCCTGCGCTTCAGCCCATGCTGCGCGCGCTGAAGCGCTTTTCGCGCTGGCTCGCTCGCT 978
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 9981 GGTTACCCTGCGCCAGCAGCTGAAAGCGCTTAAACCATTTCACTGCTTGGCGCATCGCT 10040
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 979 GCCAGTGGGTACGCGAGCGGCGCTGAGCTACTACGAGTCAACAGGCGACCAACG 1038
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 10041 TTGAAGTTGCTGGAGCATTAACGCGTACGTTGATTAACGATTCGAAGCGACCAAG 10100
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1039 TCGCGCGCGCTGGCGCGATCGAGGGCTGGGTGCGCGACATCGACGCAAGCTGGTGC 1098
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 10101 TCGGCGATGACGAAAGCGCGCTGAATGGCTG-----CACGTAGACGCGCTGCATT 10154
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1099 TGCTCGCGCGGAGACGGCGAAGGCGCGGATTTCCATGACCTGGCGGAGCGGTCGCGC 1158
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 10155 TGTGCTGGCGCGCATGTTAAATCGCGCGACTTTAGCCCACTGGCGCTTACCTGGAATG 10214
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1159 GCTTCTGCGCGGCTGCTGCTGCTGGCGCTGACGCGCGGCTGATTCGCCAGGCACTGG 1218
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 10215 GCGATAAGTACGCTCTGTTGTTTCGCTGCTGAGCGGCTGCAGCTGGCGGCGCTAC --- 10271
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1219 GCAACGGGTTACCGCTGGTGGCGCTGCAACGCTGGAGAGCAAGCTTCGCGGAGCGCGCG 1278
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 10272 ---GCCCGGAAGTGGCAGAACAAACCGAAACTATGGAACAGCGGATGCGCTTGTGGCTC 10328
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1279 AGCTGGCGCGGAGGCGATCGGTGCTGTTGTCGCGCGCTTCGCGAGACCTTGGACATGT 1338
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 10329 CGCGTGTTCAGCGCGCGATGTTGTTTACTCTCCCGCGCTGTGCCAGCTTGCATCAT 10388
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1339 TCAAGAACTTCAAGAGACGCGCGCTGTTTGGCCAAAGCGCTAGAGAGCTAGCGTGTAT 1398
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 10389 TCAAGAACTTTCAGCAACAGGCAATGATTTGCCGCTCTGGCGAGGAGTTAGTGTAT 10448
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1399 GCTGTGGTGTGGCGCCCTTCCCGCTCGCGCTGTTGAGCGGCGACGCGATC 1450
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```



```

/translacion="MKLYNLKDHNEQVSFAQAVTQGLKQKQGLFFPHDLPEFSLTEID
EMKLDFTVRSFAKILSLIGDEIPOEILERVRFAAFAPAPVANVEDSDVGCLELFGP
TLAPKDFGGRFMAQLTHIAGDKPVTILTATSGDTGAVAHAFVLPNVKVVILYPEG
KISLPQKLEFCTIGGNETETVALDGEDACALYKQAFDDDELKVALGINSANSINISR
LIAQCIYFBAVQALPOEARLNVVSPGNGDLTLAGLLAKSLGLSPVKRFIAATNVN
TDVRFHLDKNSPKAQTALNSMVDVSPNNPRVEELFRRKIWLKELGYAAVDDE
TTOQTMRELKELGYTSPPHAAVAYRALRDQNFGEYFLGTLTAHPAKFKESVAILGE
TLDLPLKAEARADLPILLSHNLPAADFAALRLKLMNHQ"
5251..5547
/gene="ECs0005"
CDS
5251..5547
/note="ECs0005"
95 in 98 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence-not_experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB33428.1"
/db_xref="GI:13359461"
/translacion="MKMQSIVLALSLLVLPAPMATAAEITLVPSVKLQIGDRNROY
YWDGHWDRDHGWKQHVYEWGRNRPWHPPPPPRHKKHAKHDDHGGHGPCKHHR"
complement(5700..6476)
/gene="ECs0006"
complement(5700..6476)
/note="ECs0006"
99 in 258 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence-not_experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB33429.1"
/db_xref="GI:13359462"
/translacion="MLLISPAKTLQSPILTRTYLPELLDNSQOLIHEARKLTPP
QISTLMRISKLGINARHDWQDPTPENAKOAILAFKGDVYTGQAEFTFSEDDFD
FAQOHLRMLSLGLVRLDPMQPYRLEMGRIENARGKLDYQFWGDIITNKLINEALA
AQEDVNVINLASDEYKSVKPKLNAEIIKPFIDEKNKFKIISFYAKKARGILMSRF
IENRLTKPQLQKGFNSGEYFFDEASSNSGELVFKRYEQR"
complement(6546..7976)
/gene="ECs0007"
complement(6546..7976)
/note="ECs0007"
99 in 476 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence-not_experimental
/transl_table=11
/product="putative inner membrane transport protein"
/protein_id="BAB33430.1"
/db_xref="GI:13359463"
/translacion="MPDFFSFINSVLGMSVMIYLLFGAGCWFTFRTGFVQFRIQFPG
KSLKNSVTHPOGGTLTSQSLCTSLAARVSGSNLAGVALAITAGGPGAVFWWVAAFIG
MAYSFACSLAOLYKERDVGQPRGGPAWYMARGLMRWGVLPFVALLIAYGIIIFSG
VQANARLASDFDFFPLVYTGIIILAVFALLIIRGLHGVARLMQGVPLMALLIWLT
SLVTCINOLCPHVIWIFESAFGQWGAAGAGYTLISQAITNGFORSFMSFEAGMG
STPNAAAASWPHPAQOIGMIGIFIDTLVICTASAMILLIAGNITGYMPLGQIC
LLOKAMVLGMSWGAEPVTLVILFAPFSIVANYIYAENNLFLRLNNKRAIWLCLRIC
TFATVIGTLLSLPLMQLADIIMACWAIITNLAILLSPVVHTIASDYLRLQRKLGR
VPFDPDIPDIGRLSRDANDVSOE"

```

Query Match  
 Best Local Similarity 53.9%; Score 288; DB 1; Length 281530;  
 Matches 734; Conservative 0; Mismatches 605; Indels 33; Gaps 5;

Qy 82 TCGTTTCGGCTCGGCAAGAGCGCATGCTCCCTGTGGCTACTCGGCGGCGCGGCT 141  
 Db 101720 TCATTATCGGCTGGCGCTCACCGGCTTTCTCGTGGACTTTTCTCGCTCGCGGTG 101779

Qy 142 TGCCTTTTCGGCTGGTATACCGAGAACCGCGGAGCTGGCCACCTCGTGGTGGCC 201  
 Db 101780 TGACGCGCGGCTTATGATACCGCTATGACACCGCTGGCTGGATAAATACCGG --- 101836

```

202 AGTATCCGAGTGGAAAGTGGTTGGCGGCAACTCGACGCGGAGTTCCTCTGCTCGCGCC 261
101837 -----AAGCGGTAGAGCCACACGCGCATCTGAATGATGAATGGCTGATGCGGCGAG 101890
262 GCGAACTCTATGTCAAGCCCGCGGTTCGTGCTGCGACACCCCTGCGCTGGTACAGCGCGCG 321
101891 ATCTGATTTGCGCCAGTCCCGGTATTGCACCTGGCGCATCCATCTTAAAGCCGCTCGCGGTG 101950
322 CGAAGGCGTGGCATCTCCGCTGACATCGATCTCTTCGCGCGGAGGCGAAGCCCGCA 381
101951 ATGCCGGAATGAAATCGTTGGCGATATCGAGCTGTCTGTGCGGAGCAACAGCCGGA 102010
382 TCGTCGCGCATCACCGGTTTCAACGCGAAGACACCGCTGACACACCTGTGTGGCGCAATGG 441
102011 TTGTGGCGATTACCGGTTCTAACGCGAAGAACACGCGTACCACGCTAGTGGTGAATGG 102070
442 CGGTGGCGGCGGACAAGGTTGTCGCTGCGCGGCAACCTCGGCAACCCCGCGGCTCGACC 501
102071 CGAAGCGCGGCGGTTTAAACGTTGGTGTGGTGGCAATATTGGCTTGCCTGCGCTGTGATGC 102130
502 TGTGCGCGGACGACATCGAGCTGTACGTTGTTGGAGCTGTGAGCTTCCAGCTGGAACCT 561
102131 TGTGATGATGAGTGGCACTGTACGTACTGGAACCTGTGAGCTTCCAGCTTGAACCA 102190
562 GCGATCGCCTCAACCGCGAGGTGGCGACCGCTGCTGAACGTGACGCAAGACCATATGGATC 621
102191 CCTCCAGCTTACAGCGGTAGCAGCACCACTTCTGAAGCTGACTGAAGATCATATGATGC 102250
622 GCTA---CGAGCGATGGCTGACTACACCTGGCGCAAGCAGCAGGATCTTCCGCGGTGCC 678
102251 GCTATCCGTTTGGTTTACGCAATACCGTGGCGCAAAATTCGCTATTATTACAAAATGCGA 102310
679 GCCAGGTCGTGGTGAATCGCGCGCATGCCCTGACCGACCGCTGATCCCGCATACCGTGC 738
102311 AAGTTTGTGCTCAATGCTGATGCTTAAACAATGCCATTCGCGGTGCGGATGAC 102370
739 CGTCTGGTCTGCTCGGCTGAACAAGCGCGACTTCAAGGCTTTCGCGCTGATCGAGGAAG 798
102371 GCTCGCTGAGCTTTGGCGTCAACATGGGTGACTATCA-----CCTGAATCATCAGC 102421
799 ACGGCCAGAGTGGCTGGCTTCCAGTTCGACAAGCTGCTGCCGGTTGGCCAACTGAGA 858
102422 AGGCGAATCTGCTCGGCTGAAAGCGAGAAAGTGTGAACGTGAAAGAAATGAAC 102481
859 TCCGTGGCGCCACCACTATTCCAAACGCGCTGCGCGCTGGCTGGCGCATCGGCTGCG 918
102482 TTTCCGGCAGCATAACTACACCAATGGCTGGCGGCTGGCTGGCAGATGCCGAG 102541
919 GCCTGCGCTTCGAGCCCATGCTCGGCGCTGGAAGGGGTTTTTCGCGCTGGCTCATCGCT 978
102542 GGTACCGCGTCCAGCAGCCCTGAAAGCGTTAAACCACTTCACTGCTGCGCGATCGCT 102601
979 GCCAGTGGTACGAGCGGCGGCGGCTGAGCTACTAGCAGATTCCAAGGCCACCAACG 1038
102602 TTGAAGTTGCTGGAGCATAACGCGTACCTTGGATTAAAGATTGCAAGCGCAACAG 102661
1039 TCGCGCGCGCTCGCGCATCAGAGGCTGGGTGCGGACATCGACGCGCAAGCTGGTGC 1098
102662 TCGCAGTACGAGCGCGCTGAATGGCTG-----CAGTAGACGCGCAGCTGCATT 102715
1099 TGCTGCGCGCGGAGAGCGCAAGGCGCGCTGATTCATGACTGCGGAGCGGCTGCGCGC 1158
102716 TGTTCGCGGCGGATGGTAAATCGCGGACTTTAGCCCACTTACCTGCTTACCTGAATG 102775
1159 GCTTCTGCCGCGGTGCTACTGCTGGCCGTGACCGCGGCTGATTCGCCAGGCACTGG 1218
102776 GCGATAACGTACGCTCTATTGTTTCGGTCTGTGAGCGGCTGCAGCTGCGCGCTAC --- 102832
1219 GCAACGCGGTACCGCTGGTGGCGCTGCAACCGCTGAGCAGCAGTCCGCGAGCGCGCG 1278
102833 ---GCCCGGAAGTGGCAGAACACCGAATCTGGAACAGCGGATGCGCTTGTGGCTC 102889
1279 AGCTGGCGCGGAGAGCGGATGCGGTGCTGTTGTCGCGCGCTGCGCGAGCGCTGGACATGT 1338

```

```

Db      102890  CGCGTGTTCAGCGCGGCGATATGGTTCTACTCTCCCGCGCTGTGCCAGCCTTGATCACT 102949
QY      1339    TCAGAAGACTTCGAAGAACCGGACGCGCTGTTCGCCAAAGCCGAGAGGAGCTAGCGTGAT 1398
Db      102950  TCAGAAGACTTCGAACACAGGCAAGTACGATTTGCCCGCTGTGGCGAAGGAGTTAGGTTGAT 103009
QY      1399    GCTGTCGGTGTTCGCGCCCTTCCCGTCGCCGTGTGAGCCGCGACGGCATC 1450
Db      103010  GCCTTTATCTCTCCCTCGCTGAATAATGCCGCGCTGCCAGGATTCAGTATC 103061

RESULT 12
LOCUS   AE016755                      300409 bp    DNA    linear    BCT 24-DEC-2002
DEFINITION Escherichia coli CFT073 section 1 of 18 of the complete genome.
ACCESSION AE016755 AE014075
VERSION   AE016755.1 GI:26106314
KEYWORDS
SOURCE   Escherichia coli CFT073
ORGANISM Escherichia coli CFT073
          Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
          Enterobacteriaceae; Escherichia.
REFERENCE
AUTHORS  Welch,R.A., Burland,V., Plunkett,G.D. III, Redford,P., Roesch,P.,
          Rasko,D.A., Buckles,E.L., Liou,S.-R., Boutin,A., Hackett,J.,
          Stroud,D., Mayhew,G.F., Rose,D.J., Zhou,S., Schwartz,D.C.,
          Perna,N.T., Mobley,H.L.T., H.L.T., Donnenberg,M.S. and Blattner,F.R.
          Extensive Mosaic Structure Revealed by the Complete Genome Sequence
          of Uropathogenic Escherichia coli
JOURNAL  Proc. Natl. Acad. Sci. U.S.A. 99 (26), 17020-17024 (2002)
PUBMED   12471157
REFERENCE
AUTHORS  Welch,R.A., Burland,V., Plunkett,G.D. III, Redford,P., Roesch,P.,
          Rasko,D.A., Buckles,E.L., Liou,S.-R., Boutin,A., Hackett,J.,
          Stroud,D., Mayhew,G.F., Rose,D.J., Zhou,S., Schwartz,D.C.,
          Perna,N.T., Mobley,H.L.T., H.L.T., Donnenberg,M.S. and Blattner,F.R.
          Direct Submission
JOURNAL  Submitted (20-JUN-2002) Genetics Laboratory, University of
          Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA
FEATURES
          Location/Qualifiers
          1..300409
             /organism="Escherichia coli CFT073"
             /mol_type="genomic DNA"
             /strain="CFT073"
             /db_xref="taxon:199310"
             190..255
             /gene="thrL"
             /locus_tag="c5491"
             190..255
             /gene="thrL"
             /locus_tag="c5491"
             /function="leader"
             /note="Thr operon attenuator; Escherichia coli K-12
             ortholog: b0001; Escherichia coli O157:H7 ortholog: z0001"
             /codon_start=1
             /transl_table=11
             /product="Thr operon leader peptide"
             /protein_id="AA078501.1"
             /db_xref="GI:26106315"
             /translation="MKRISTTITTTITTTTGTGNGAG"
             308..1018
             /insertion_seq="IS1541A-like"
             534..911
             /locus_tag="c0002"
             534..911
             /locus_tag="c0002"
             /note="Residues 2 to 125 of 125 are 94.35 pct identical to
             residues 29 to 152 of 152 from SwissProt.40 :
             >sp|Q57334|T200_SALTY Transposase for insertion sequence
             element IS200"
             /codon_start=1
             /transl_table=11

```

```

/product="Hypothetical protein"
/protein_id="AA078502.1"
/db_xref="GI:26106316"
/translation="MFYREKRAIGCILKLCWKSVIRLEAECCADHIMLVETPPK
MSVSGFMGLKSKSLMPVEQFDLKFYRNREFWCRGYVVDTVGKNTAKIQDIYIKHQ
LEEDKMGEOGLSPYFGSPFTGRK"
985..3510
/gene="thrA"
/locus_tag="c0003"
985..3510
/gene="thrA"
/locus_tag="c0003"
/function="enzyme; Amino acid biosynthesis: Threonine"
/note="Bifunctional; Escherichia coli K-12 ortholog:
b0002; Escherichia coli O157:H7 ortholog: z0002;
bifunctional; Homoserine dehydrogenase I"
/codon_start=1
/transl_table=11
/product="Aspartokinase I"
/protein_id="AA078503.1"
/db_xref="GI:26106317"
/translation="MKNRLCRRIFTSTKGNVTTMRVLKFGSTSVANAERFLRVADI
LESNARQGVATVLSAPAKITNHLVAMIEKTISSGODALPNISDAERIFAEILTLGAAA
QGFPLAQLTKTVDQEFFAOIKHVLHGSIAGOCPSINAALICRGKEMSIATMAGVLE
ARGHNVTIDPVEKLLVGHYLESTVIAESTRTTAASRIADHMHVLMAGFTAGNEKG
ELSVLRGNSDYSAVALAACLRADCEIWTVDGYTCDPQVQPARLLKSKSYQFAM
ELSYFGAKVLHPRTTTPIAQFOIPLIKNTGPNQAPGTIGASRDEDELPKVSGISNLN
NAMPSVSGPMGKMGVMAARVFAAMSRARISVLLITSSSEYSISFCVQSDCVRAE
RAMQEFLELEKGLLEPLAVTERLAISVVGDMRTLRGISAKFPAALARANINVA
IAQSSERSISVNNDDATTGVTTHQMLFNTDQTVIEVFYIGVGVGALLDEQLKRA
QSMLNKHIDLRVCGVANSKALLTNHGLNLENWQELAAQAKPEFNLRLILRVKEYH
LINPVIVDCTSSQAVDQYADFREGFHVVTNNKANTSSMDYIHLRYAABKSRKE
EMGYTEPPRODLSMDVARKLLILARETGRELEADTEIEPVLPAEFNAEGDVAEFM
LVYTNVAGLPIVENLQNLNAGDELKPSGLSLSYIFGKLDGMSFSFSAATTLAR
ANLSQDLNLFARVAKARDEGKLVRYVGNIDEDGVCVRVKAIEVDSNDPLFKVKNGENA
LAFYSHYQPLPLVLRGAGNDVTAAGVFADLLRLTSLWKLCV"
3512..4444
/gene="thrB"
/locus_tag="c0004"
3512..4444
/gene="thrB"
/locus_tag="c0004"
/function="enzyme; Amino acid biosynthesis: Threonine"
/note="Escherichia coli K-12 ortholog: b0003; Escherichia
coli O157:H7 ortholog: z0003"
/codon_start=1
/transl_table=11
/product="Homoserine kinase"
/protein_id="AA078504.1"
/db_xref="GI:26106318"
/translation="MVKVYAPASSANMSVGFDLGAAVTPVDGALLGDVVTVEAAETF
SINNLRPADKLPSPRENIVYQCWERFCQELGKOIPVAMTLEKNMPISGSGIGSSACS
VVAALMANNEHCCKPLNDRLLALMGELEGRISGSIHDVNPVAPCLGMLQNIENDI
ISQQVPGDELDWLVAIPGKIVSTAEARAILPAQRQDCIAGHGRFTIACISRSRQ
PELANMLMDVIAEYPERLLPFGFARQAEVGAIEGAVASGISGSGPTFALCDKPDPTA
QRVADWLGNLYLQNGEFGVHCIRLDTAGARVLEN"
4445..5731
/gene="thrC"
/locus_tag="c0005"
4445..5731
/gene="thrC"
/locus_tag="c0005"
/function="enzyme; Amino acid biosynthesis: Threonine"
/note="Escherichia coli K-12 ortholog: b0004; Escherichia
coli O157:H7 ortholog: z0004"
/codon_start=1
/transl_table=11
/product="Threonine synthase"
/protein_id="AA078505.1"
/db_xref="GI:26106319"
/translation="MKLYNLDHNEQVSFAQAVTQGLGKNQGLFFPHDLPFSLTEID

```





Db 101996 TGCGATPAACGTACGTGTATTGTTTCGGTCGTGACGGCGCGAGCTGGCGCGCTAC- 102054  
 QY 1217 GGGCAACGCGGTACCGTGTGGCGGCGCAACGCTGAGCAAGCAGTCCGGCAGGCCG 1276  
 Db 102055 -----GCCCGAGTGCAGAACAAACCGAAACTATGGAACAGCGATGCGCTGTGCTTC 102109  
 QY 1277 CGAGCTGGCCCGGAGCGGATGCGGTGTGTTGTCGCCGCGCTGCGGAGCCTGGACAT 1336  
 Db 102110 TACGCGTGTTCAGCGCGGATATGTTGTCTCTCCCGCGCTGTGCCAGCCTTGATCA 102169  
 QY 1337 GTTCAAGAACTTCGAAGACCGGACCGCTGTTCGCCAAAGCCGTAGAGAGTAGCGTG 1396  
 Db 102170 ATTCAGAACTTGAACACAGGCAATGAGTTTGCCCGCTGTCGCGAAGGAGTAGTTG 102229  
 QY 1397 ATGC 1400  
 Db 102230 ATGC 102233

## RESULT 13

LOCUS ECMUROY 2608 bp DNA linear BCT 12-SEP-1993  
 DEFINITION Escherichia coli murd gene and ORF-Y (EC 6.3.2.9).  
 ACCESSION X51584.X52117  
 VERSION X51584.1 GI:42058  
 KEYWORDS membrane protein; murd gene;  
 UDP-N-acetylmuramoyl-L-alanyl-D-glutamate synthetase; unidentified  
 reading frame.

## SOURCE

ORGANISM Escherichia coli  
 Escherichia coli

Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 Enterobacteriaceae; Escherichia.

## REFERENCE

AUTHORS Ikeda,M., Wachi,M., Ishino,F. and Matsubashi,M.  
 TITLE Nucleotide sequence involving murd and an open reading frame ORF-Y  
 spacing mufF and ftsW in Escherichia coli  
 JOURNAL Nucleic Acids Res. 18 (4), 1058 (1990)

## MEDLINE

90192099  
 PUBMED 2179861

## REFERENCE

2 (bases 1 to 2608)  
 Ikeda,M.  
 Direct Submission  
 Submitted (02-JAN-1989) Ikeda M., Institute of Applied  
 Microbiology, The University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku,  
 Tokyo 113, Japan  
 3 (bases 26 to 1354)  
 Flouret,B.

## AUTHORS

Flouret,B.  
 Submitted (15-MAR-1990) Flouret B., URA 1131 du Centre National de  
 la Recherche Scientifique, C N R S, Biochimie Moleculaire et  
 Cellulaire, Batiment 432 Universite Paris-Sud, 91405 Orsay, France  
 Sequence independently determined by [3], but authors accept  
 sequence [1] as definitive.

## FEATURES

## source

1..2608  
 /organism="Escherichia coli"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:562"  
 /map="2min"  
 98..102  
 /note="pot. ribosome binding site"

## RBS

111..1193  
 /note="ORF-Y (AA 1-360)"

## CDS

/codon\_start=1  
 /transl\_table=11  
 /protein\_id="CAA35932.1"  
 /db\_xref="GI:42059"  
 /db\_xref="SWISS-PROT:P15876"  
 /translation="MDVLAELHVKYISGFVNSYLTFRAYISLLTALFISLMGPRM  
 IAHQLSFGVVRNDPESHFSGRTPTMGIMILTAIVISVLLWAYSPNVPVCL  
 VVLVGVTGFDVDDYKVVYKDTKGLIARWKYFVMSVIALGVAFALYLAGKDTPATQL  
 VVPFKDVMPLGLFILLAYFVIVGTGNVNLTDGLDGLAIMPVFAAGPALVAA

TGMNFASYLHPIPLRHAGELIVICTAIVGAGLFLWFNTYPAQVFMGDVCSLALGGA  
 LGTIIIVRREFLLVIMGVVYVETLSVLQVGSFKLRGQIRFIRMAPIHIIHYELKGNP  
 EPVIVRFFIISLMLVLIGLATLKR"  
 1018..1023  
 /note="pot. -35 region"  
 1043..1048  
 /note="pot. -10 region"  
 1181..1185  
 /note="pot. ribosome binding site"  
 1196..2512  
 /note="murd gene product (AA 1-438)"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="CAA35933.1"  
 /db\_xref="GI:42060"  
 /db\_xref="SWISS-PROT:P14900"  
 /translation="MADYQGNKVVIIIGLTLGSLCVDFFLARGVTRVMDTRMTPPGL  
 DKLPEAVERHTGSLNDEWLMMAADLIIVASGIALAHPSLSAADAGIEIVDIEFCRE  
 AQAPIVATITSGNGKSTVTLVGEAKAAGVNVGVGNIGLIPALMLLDCELELYLELS  
 SFOLETTSLQAAVATILNVTEHMDRYPEGLQYRAAKLRIVENAKVCVVNADALT  
 MPTRGADERCVSFGVNMGDYHLNHQGETWLVKGEKLVNVEKLSGHOHNTNALAA  
 LALADAAGLPRASSLSKALTFTGLPHRFEVVLHNGVWINDSKATNVGSTEALNGL  
 HVDGTHLLLGSGKADSPFLARYLNGDNVRLYCFGRDGAQAALRPEVAEQTETME  
 QAMRLIAPRVQDGMVLLSPACASLDQFKNFQGRNEFARLAKELG"  
 BASE COUNT 530 a 631 c 777 g 670 t

Query Match 19.4%; Score 282; DB 1; Length 2608;

Best Local Similarity 53.2%; Pred. No. 1.1e-24;  
 Matches 731; Conservative 0; Mismatches 610; Indels 33; Gaps 5;

QY 80 CATCGTTGTCGGCTCGGCAAGAGCGGCATGTCCCTGGTGCCTACCTGGCGGCCCGCG 139  
 Db 1222 CGTCATTATCGGCTGGCGCTCACCGGCTTCTCGTGGACTTTTCTCTCGCTCGCG 1281  
 QY 140 CTTGCGCTTTCGGCGTGTGTGATACCGGAGAGAACCGCCGCGAGCTGCCACCTCGGTGC 199  
 Db 1282 TGTGACGCCGCGCGTGTATGATACGCGTATGACACCGCTTGGCTGGATAAATACCCG- 1340  
 QY 200 CCAGTATCGCAGGTGGAAGTGGCTTGGCGGCAACTCGACGCGGAGTTCCTCTGCTCCGC 259  
 Db 1341 -----AAGCGGTAGAACGCCACACGGGCAGTCTGATGATGATGGCTGATGGCGG 1392  
 QY 260 CCGCGAACTCTATGTGACGCCCCGGCTGTGCTGCGCACCCCTGCGCTGTGTACAGGCCG 319  
 Db 1393 AGATCTGATGTGCGCAGTCCCGGATTTGCACTTGGCGCATCCATCTTAAAGCGCTGCCG 1452  
 QY 320 CGCGAAAGGCGTCCGCTATCTCCGTCACATCGATCTTCCGCCCGGAGCGGAGGCCCC 379  
 Db 1453 TGATGCCGAATCGAAATCGTTGGCGATATCGAGCTGTCTGTGCGGAAGCACAAGCACC 1512  
 QY 380 GATCGTCCGCTATCACCAGTTTCCAAACGGAAGAGCAGCGTGAACACCTTGGTGGCGAAAT 439  
 Db 1513 GATTGTGGCGATTACCGGTTCTAACGCGCAAAACAGCGTCAACACGCTAGTGGGTGAAT 1572  
 QY 440 GCGGCTGGCCGCGGACAAGCGTGTCCGCTGCGGCGCAACCTCGGCACCCCGCGGTGCA 499  
 Db 1573 GCGCAAGCGCGCGGGTTAAACCTTGGTGTGGGCAATATTTGGCTGCTGCTGCTGAT 1632  
 QY 500 CCTGCTGGCCGACGACATCGAGCTGTACGTGTGGAGCTGTCCAGCTTCCAGCTGGAAAC 559  
 Db 1633 GCTACTGGATGATGAGTGTGAAGTGTACGCTGGAACTGTGCGAGTTCGAGCTTCAGCTG 1692  
 QY 560 CTCGATCGCTCAACCGCGAGGTGGCGAGCTGTGAAGCTGAGCGGAAGACCATATGA 619  
 Db 1693 CACCTCCAGCTTACAGCGGTAGCAGGACCATCTTGAAGCTGACTGAAGATCATATGA 1752  
 QY 620 TCGCTA---CGACGCGATGGCTGACTACCACTGGCCAAAGACCGACCTTCCTCCGCGTGC 676  
 Db 1753 TCGCTATCGGTTTGGTTTACAACAGTATCGTCAGCAAAAACCTCGCATTTTACGAAACGC 1812  
 QY 677 CCGCCAGGTGCGGTGAATCGGCCCATGCCCTGACCCCGCCGCTGATCGCCGATACCTG 736



```

Db      1813 GAAAGTTTCGCTGTTAATGCTGATGATGCCCTTAACAATGCCGATTCGCCGTGCGGATGA 1872
QY      737  GCGGTGCTGGTTCGCTGGCTGAACAAGCCGAGCTTCAAGGCTTTTCGCCCTGATCGAGGA 796
Db      1873 ACGTGCCTCAGCTTTGGCGTCAACATGGGTGACTATCA-----CCTGAATCATCA 1923
QY      797  AGACGGCCAGAGTGGCTGGCGTTCCAGTTCGACAAGCTGTCCGGTGTGGCGAACTGAA 856
Db      1934 GCAGGGCAAACTGGCTGGCGGTTAAAGGCGAGAAAGTGTGAATGTGAAAGAGATGAA 1983
QY      857  GATCGTGGCGCCACACATATTCCACGCGCTGCCCGCTGGCGCTGGCGCCATCGCGGT 916
Db      1984 ACTTCCGGGCGAGCATACACCAATCGCTGGCGCGCTGGCGCTGGCGAGATGCTGC 2043
QY      917  CGGCTGCGCTTCGACGCCATATTCGCGCGCGCTGAAGCGCTTTCCCGCCCTGGCTCAFCG 976
Db      2044 AGGTTTACCGGTGCCAGCAGCCCTGAAGCGCTTAACCAATCACTGCTGCTGCCGCATCG 2103
QY      977  CTGCCAGTGGTACGAGCGGAGCGGTGAGCTGAGCTACTACACGATTCACAGGCCACCAA 1036
Db      2104 CTTTGAAGTTGTCTGGAGCATAACGCGGTAGCTTGGATTAACTGATTCGAAAGGACCAA 2163
QY      1037 CGTGGCGCCGCTTCGCGCGATCGAGGGCTGGTGCACATCGACGCGCAAGCTGGT 1096
Db      2164 CGTCGGCAGTACGGAAGCGGCTGAATGGCTG-----CACGTAGACGCACACTGCA 2217
QY      1097 GCTGCTCGCGCGGAGACGCAAGGCGCGGATTCATCACTGACCTGCGCGAGCGGTGCG 1156
Db      2218 TTTTGTGCTGGTGGCATGTAATCGCGGACTTTAGCCCACTGGCGGCTTACCTGAA 2277
QY      1157 GCGCTTTCGCGGGGTGTTACTGCTTGGCGGTGACCGCGGCTGATTCGCCAGGCACT 1216
Db      2278 TGGCGATAACACTAGCTGTGATTGTTTGGTTCGTGACGCGCGCGAGCTGGCGGCGTAC- 2336
QY      1217 GGGCAACGCGTACGCTGGTGGTGGCGTGCACACCTGGAGCAAGCAGTCCGCGAGCGCG 1276
Db      2337 -----GCCGGAATGGCAGAACAAACGAACTATGGAACAGCGGATGCCCTTGTGGC 2391
QY      1277 CGAGCTGGCGCGGAGGCGATGCGTGTCTGTCGCGGCTTCGCGAGCCTGGACAT 1336
Db      2392 TCCGCTGTTTCAGCGCGGATATGTTCTGCTCTCCCGACGCTGTGCCAGCCTTGATCA 2451
QY      1337 GTTCAAGAACTTCAAGAACGCGAGCGCTGTTGCGCAAGCCGTAGAGGAGTAGCGTG 1396
Db      2452 GTTCAAGAACTTTGAACAACGAGCAATGAGTTTGCCTGCTGCGCGAGGAGTTAGTTG 2511
QY      1397 ATGCTGTCGGTGTGGCGCCCTTCCCGTGGCGCTGTTGAGCGCGCAGCGCATC 1450
Db      2512 ATGCGTTTACTCTCCCTCGCTGAAATGCCGCGCTGCCAGGATTCAGTATC 2565

RESULT 14
LOCUS   AE000118
DEFINITION Escherichia coli K12 MG1655 section 8 of 400 of the complete genome.
VERSION AE000118.1 U00096
KEYWORDS AE000118.1 GI:1786262
SOURCE  Escherichia coli K12
ORGANISM Escherichia coli K12
REFERENCE Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
          Enterobacteriaceae; Escherichia.
AUTHORS Blattner, F.R., Plunkett, G. III, Bloch, C.A., Perna, N.T., Burland, V.,
          Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F.,
          Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J.,
          Mau, B. and Shao, Y.
TITLE    The complete genome sequence of Escherichia coli K-12
JOURNAL Science 277 (5331), 1453-1474 (1997)
MEDLINE 97426617
PUBMED  9278503
REFERENCE 2 (bases 1 to 21757)

```

## AUTHORS

Blattner, F.R.  
Direct Submission  
Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,  
University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.  
Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax:  
608-263-7459

## REFERENCE

3 (bases 1 to 21757)  
Blattner, F.R.  
Direct Submission  
Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,  
University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.  
Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax:  
608-263-7459

## REFERENCE

4 (bases 1 to 21757)  
Plunkett, G. III.  
Direct Submission  
Submitted (13-OCT-1998) Laboratory of Genetics, University of  
Wisconsin, 445 Henry Mall, Madison, WI 53706, USA

## COMMENT

This sequence was determined by the E. coli Genome Project at the University of Wisconsin-Madison (Frederick R. Blattner, director). Supported by NIH grants HG00301 and HG01428 (from the Human Genome Project and NCHGR). The entire sequence was independently determined from E. coli K12 strain MG1655. Predicted open reading frames were determined using Genemark software, kindly supplied by Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA, 30332 [e-mail: mark@ember.gatech.edu]. Open reading frames that have been correlated with genetic loci are being annotated with CG Site Nos., unique ID nos. for the genes in the E. coli Genetic Stock Center (CGSC) database at Yale University, kindly supplied by Mary Berlyn. A public version of the database is accessible (<http://cgsc.biology.yale.edu>). Annotation of the genome is an ongoing task whose goal is to make the genome sequence more useful by correlating it with other data. Comments to the authors are appreciated. Updated information will be available at the E. coli Genome Project's World Wide Web site (<http://www.genetics.wisc.edu>). \*\*\* The E. coli K12 sequence and its annotations are periodically updated; this is version M34. No sequence changes. Annotation updates: updated gene identifications and products; all new functional assignments courtesy of Monica Riley; added promoters, protein binding sites, and repeated sequences described in reference 1. The unique numeric identifiers beginning with a lowercase 'b' assigned to each gene (protein- or RNA-encoding) are now designated as gene synonyms instead of labels. This should allow them to be searched for in Entrez as gene names.

## FEATURES

Location/Qualifiers  
1. 21757  
/organism="Escherichia coli K12"  
/mol\_type="genomic DNA"  
/strain="K12"  
/sub\_strain="MG1655"  
/db\_xref="taxon:8333"  
complement(77..163)  
/gene="leuL"  
/note="synonym: b0075"  
complement(77..163)  
/gene="leuL"  
/function="leader; Amino acid biosynthesis: Leucine"  
/note="f28; 100 pct identical to LPL\_ECOLI SW: P09149"  
/codon\_start=1  
/transl\_table=11  
/product="leu operon leader peptide"  
/protein\_id="AAC73186.1"  
/db\_xref="GI:1786263"  
/translation="MTHIVRFIGILLNASSLRGRRVSGIOH"  
complement(197..226)  
/note="factor Sigma70; promoter leu; documented +1 at 83735"  
complement(201..221)  
/note="No predicted promoter"  
/bound\_moiety="LexA predicted site"  
complement(201..221)  
/note="central position to predicted promoter: -0.5"

## gene

source

## CDS

gene

## promoter

protein\_bind

## protein\_bind

protein\_bind

```

/bound_moiety="LexA predicted site"
complement(354..368)
/notes="No predicted promoter"
/bound_moiety="Lrp documented site"
complement(354..368)
/notes="central position to leuABCD promoter: -156.5"
460..474
/bound_moiety="Lrp documented site"
/notes="central position to predicted promoter: -28.5"
462..490
/notes="factor Sigma70; predicted +1 start at 84042"
complement(519..537)
/notes="central position to predicted promoter: -319.5"
554..565
/notes="tyrR predicted site"
/notes="central position to predicted promoter:64"
/bound_moiety="AraC predicted site"
646..1767
/genes="leuO"
/notes="synonym: b0076"
646..1767
/genes="leuO"
/function="putative regulator; Amino acid biosynthesis:
Leucine"
/notes="o373; 100 pct identical to LEUO_ECOLI SW:
P10151(290 aa) but contains 59 additional N-ter aa and 24
C-term residues; 100 pct identical to PIR: S40589 but
contains 24 additional C-term residues"
/codon_start=1
/transl_table=11
/product="probable transcriptional activator for leuABCD
operon"
/protein_id="AAC73187.1"
/db_xref="GI:1786264"
/transl_table=11
/translation="MPHSTAMDSVFTRIFEMSEYSCFFLFYMHDKSYSGGLFLC
IPRRELSVTYVLSMPVQTDHPETAELSKPOLRMVDNLNLTVDFAVQEQNITRAA
HVLGMSQPAVSNARVAKVFNDELFVYRGIGQPTARAFQFGSVRQALQVONELP
GSGFEFASPERVHLKVGSLDLSITQIYNHIEQIAPNIVHMFKSLNQNTQHOHLYR
QETFFVISEDHREFTSVPLFKDEMVLVAKNHPTIRGPLLKHVYNEQHAASLD
RFASFQPMYDITVDKQASITAYQGMAMSVLSVVSQTHLVAIPRWLAEEFAESLELQV
LPLPLKQNSRTCYLSHHEAAGRDKGHWEEQLVLSICKR"
1802..1816
/notes="central position to predicted promoter: -39.5"
/bound_moiety="Lrp predicted site"
1813..1839
/notes="factor Sigma70; promoter ilvIHp4; documented+1 at
85394"
1833..1847
/notes="central position to predicted promoter: -39.5"
/bound_moiety="Lrp predicted site"
1838..1866
/notes="factor Sigma70; promoter ilvIHp3; documented+1 at
85420"
1915..1929
/notes="central position to predicted promoter:73.5"
/bound_moiety="Lrp predicted site"
1949..1963
/notes="central position to predicted promoter:107.5"
/bound_moiety="Lrp predicted site"
1955..1984
/notes="factor Sigma70; promoter ilvIHp2; documented+1 at
85534"
1978..1992
/notes="central position to predicted promoter:136.5"
/bound_moiety="Lrp predicted site"
1995..3809
/genes="ilvI"
/notes="synonym: b0077"
1995..3809
/genes="ilvI"
/EC_number="4.1.3.18"
/function="enzyme; Amino acid biosynthesis: Isoleucine,

```

```

Valine"
/notes="o604; 98 pct identical (1 gap) to 522 residues
from ILV1_ECOLI SW: P00893 (566 aa)"
/codon_start=1
/transl_table=11
/product="acetolactate synthase III, valine sensitive,
large subunit"
/protein_id="AAC73188.1"
/db_xref="GI:1786265"
/transl_table=11
/translation="MPMSLAANCLSKIGRLKMCFTHPROTVRQAMHLSGAEMVVRSL
IDGVKQVFGPGAVLDIYDALHTVGGIDHVRHEQAHHADGJARATGEVGVVL
VTSGGATNAITGIATAYMDSIPLVLSGVATSLIGYDAFQCDMVGISRPVVKHSF
LVQGTEDIPQVLKFAWLAASGRPGVVDLPKIDILNPANKLPYVNPESVMSRYNPT
TTGHQGIKRALQTLVAAKPPVYVGGAITAGCQOLKETETALNLPVVCVSLMGLGA
FPATHRQALGMLGMHGTYEANMTMHADVIFAVGVFRDRTTNNAKYCPNATVLIHD
IDPTSIKVTADIPVGDAROVLEOMLELSOESAHOPLDEIRDMQOIEQARBOC
LYDTHSEKIKPQAVIETLRLFKGDAYTSDVQGHQMFALYFPDKPRWLNNSGL
GTMFGFLPALGVKMALEPVTVCVTGDSIQMNIQELSTALQIETPLVVLNNRYL
GMVKQWDMYISGRHSOSYMSLOSLPFLAEAYGHVGIQISHPHELESKLEALEQVR
NNRLVFVDVTVDGSEHVYPMQIRGGGMDENWLSKTERT"
1998..2012
/genes="ilvI"
/notes="central position to predicted promoter:156.5"
/bound_moiety="Lrp predicted site"
2017..2045
/genes="ilvI"
/notes="factor Sigma70; promoter ilvIHp1; documented+1 at
85597"
3812..4303
/genes="ilvH"
/notes="synonym: b0078"
3812..4303
/genes="ilvH"
/EC_number="4.1.3.18"

Query Match      19.4%; Score 282; DB 1; Length 21757;
Best Local Similarity 53.2%; Pred. No. 6.6e-25;
Matches 731; Conservative 0; Mismatches 610; Indels 33; Gaps 5;

Qy 80 CATCGTTGTGGCGCTCGGCAAGAGCGGCTGTCCTGGTGGCTACCTGGCGCGCGG 139
Db 13568 CGTCATTATCGGCTGGCGCTCACCGGGCTTCTCGGTGGAGCTTTTCTCGCGG 13627

Qy 140 CTTGCGCTTTCGCGGTGCGATACCCGAGAGAACCCGCGGAGCTGCCACCTGCGTGC 199
Db 13628 TGTGACGCGCGCGGTATGATGATACCGGTATGACACCGCTTGGCTGATAATACCG- 13686

Qy 200 CCAGTATCCGACGAGTGGCTTGGCGGGAACCTCGACCGCGAGTTCCTCTGCTCCG 259
Db 13687 -----AAGCGGTAGAACGCCACACGCGGCACTCTGAATGATGGCTGATGGCGGC 13738

Qy 260 CCGCGAATCTATGTGTCAGCCCGGCTTGTGCTGCGCACCCCTGCGCTGTACAGGCGCG 319
Db 13739 AGATCTGATGTGCGCAGTCCCGGTATGACCTGGCGCATCATCTTAAGCGCTGCCG 13798

Qy 320 CGGAAAGCGGTGCGCATCTCCGGTGACATCGATCTCTTTCGCCCGGAGCGAAGGCC 379
Db 13799 TGATGCGGAATCGAAATCGTTGGCGATATCGAGCTGTTCTGTCGGGAAGCACAAGC 13858

Qy 380 GATCGTCGCATCACCGGTTCGAACCGGAGACCGGTGACCACTGCTGTCGGCGGAAT 439
Db 13859 GATTGTGGCGATTACCGGTTCTAACGGCAAAACACGGTCAACCACTAGTGGGTGAA 13918

Qy 440 GCGGTGGCGCGGACAAAGCGTGTGCGCGTTCGCGGCAACCTCGCACCCCGCGCTCG 499
Db 13919 GCGAAAGCGCGGGGGTTAACGTTGTTGGGTGGCAATATTTGGCTCGCTCGTTGAT 13978

Qy 500 CTTGTCGCCGACGACATCGAGCTGACGTGTGTTGGAGCTGTCGAGCTTCAGCTGAAAC 559
Db 13979 GCTACTGGATGATGAGTGTGAACGTGACGTGCTGGAACGTGTCAGCTTCAGCTGAAAC 14038

Qy 560 CTGCGATCCCTCAACGCGGAGGTGGCGACCGTGTGTAACGTCAGCGAGACCATATGGA 619

```

Db 14039 CACCTCCAGCTTACAGCGGTAGCAGCGACCAATCTTGAACGTACTGAAGATCATATGGA 14098  
Qy 620 TCGCTA---CGACGCGATGGCTGACTACCACTGCGCAAGCAGCGATCTCCCGGTGC 676  
Db 14099 TCGCTATCCGTTGGTTTACACAGTATGTCGACGAAATCTGCGCATTTACGAAACGC 14158  
Qy 677 CCGCCAGGTGCTGTGATCGCGCGGATGCCCTGACCGACGCGTGTATCCGCGCATACCGT 736  
Db 14159 GAAAGTTTGGTGGTTAATGCTGATGATGCTTAACAATCCGATTCGCGGTGGGATGA 14218  
Qy 737 GCGGTGCTGGTGGCTGCGCTGAAACAGCGGAGCTTCAAGGCTTTCCGCGCTGATCGAGGA 796  
Db 14219 ACGTGGCTGAGCTTTGGCTGCAACATGGGTGATATCA-----CCTGAATATCA 14269  
Qy 797 AGACGGCAGAGGTGCTGGCTTCCAGTTTCGACAACTGCTGCCGTTGGCGAACTGAA 856  
Db 14270 GCAGGGCAACCTTGGCTGGGGTTAAGCGGAGAAAGTCTGAATGTGAAGAGATGAA 14329  
Qy 857 GATCCGTGGGCGCCACAACTATTCCAAACGGCTCGCGCTGGCGTGGGCCATGCGGT 916  
Db 14330 ACTTTCGGCGCAGCATAACTACACCAATGCGTGGCGCGCTGGCGCTGGCAGATGCTG 14389  
Qy 917 CGGCTCGCTGACGCCCATGCTCGCGGCTGCAAGCGGTTTCCGGCTGGCTCATCG 976  
Db 14390 AGGTTTACCGCTGCCAGCAGCCTTGAAGCGTTAACCACTTCACTGGTCTGCCGCTCG 14449  
Qy 977 CTGCCAGTGGTACGCGAGCGGCGAGGCGGTGAGCTACTACGACGATTCCAAAGGCCACAA 1036  
Db 14450 CTTTGAAGTTGCTGGAGCATACGCGGTACGTTGATTAACGATTTCGAAAGCGACAA 14509  
Qy 1037 CGTCCGCGCGCCCTGCGCGGATTCGAGGGGCTGGTTCGCGACATCGACGCAAGCTGGT 1096  
Db 14510 CGTCCGCGAGTACGGAAGCGCGCTGAATGGCTG-----CACGTAGACGCGACACTGCA 14563  
Qy 1097 GCTCTCGCGCGGAGACGCGAGCGGCGGATTTCCATGACCTGCGCGAGCGGCTGCG 1156  
Db 14564 TTTGTTGCTGGGTGGCGATGATAAATCGCGGACTTTAGCCCACTGCGCGGTTACCTGAA 14623  
Qy 1157 GCGCTTTCGCGCGGTGGTACTGCTTGGCGGTGACGCGCGGCTGATTCGCCAGGCACT 1216  
Db 14624 TGGCGATAACGTACGTCTGATTGTTTCGTTGCTGACGCGCGAGCTGGCGGCTAC- 14682  
Qy 1217 GGGCAACGGGTACCGTGGTGGCGCTGCGCAACGCTGGAGCAAGCATTCGCGGAGCGCG 1276  
Db 14683 -----GCCCGAGTGGCAGAACAAACCGAACTATGGAACAGCGCATGGCTTGCCTGGC 14737  
Qy 1277 CGAGCTGGCGCGGAGCGCATGGGTGCTGTTGCTGCGCGGCTGCGCGAGCGCTGGACAT 1336  
Db 14738 TCCCGGTTCAGCGCGCGATATGGTTCGCTCTCCCGAGCCTGTGCCAGCCTTGATCA 14797  
Qy 1337 GTTCAAGAACTTCGAAGACGCGGAGCGCTGTTCCGCAAGCCGTAGAGAGCTAGCGTG 1396  
Db 14798 GTTCAAGAACTTTGAACAACAGGCAATGATTTGCCGCTGCGGAGGAGTTAGTTG 14857  
Qy 1397 ATGCTGCTGGTGGCGCGCTTCCCGTGGCGCTGTTGAGCGCGGACGCGATC 1450  
Db 14858 ATGCGTTTATCTCTCCCTCGCCTGAAATGCGCGCGCTGCGCAGGATTCAGTATC 14911

RESULT 15  
AX191720 AX191720 28277 bp DNA linear PAT 15-AUG-2001  
DEFINITION Sequence 2 from Patent WO0149775.  
ACCESSION AX191720  
VERSION AX191720.1 GI:15209889  
KEYWORDS Escherichia coli  
SOURCE Escherichia coli  
ORGANISM Escherichia coli  
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
Enterobacteriaceae; Escherichia.  
REFERENCE 1  
AUTHORS Iversen, P.L.  
TITLE Antisense antibacterial cell division composition and method

JOURNAL Patent: WO 0149775-A 2 12-JUL-2001;  
Avi Biopharma, Inc. (US)  
FEATURES Location/Qualifiers  
source 1. 28277  
/organism="Escherichia coli"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:562"  
BASE COUNT 6714 a 6725 c 7993 g 6845 t  
ORIGIN  
Query Match 19.4%; Score 282; DB 6; Length 28277;  
Best Local Similarity 53.2%; Pred. No. 6.2e-25;  
Matches 731; Conservative 0; Mismatches 610; Indels 33; Gaps 5;  
Qy 80 CATCGTTTTCGCGCTCGCAGAGCGGCATCTCCCTGGTGGCTACCTGGCGCGCGCGG 139  
Db 13643 CGTCATTATCGCTGGCGCTCACCGGCTTCTCTGGTGGACTTTTCTCTCGCTCGCGG 13702  
Qy 140 CTTGCCCTTTCGCGCTGCTGATACCCGAGAGAACCCGCGGAGCTGGCGACCCCTCGCTGC 199  
Db 13703 TGTGACCGCGCGCTTATGGATACGCGTATGACACCGCTGGCTGGATTAATACCG- 13761  
Qy 200 CCACTATCCCGAGGTGAAGTGGTGGGGAACTGAGCCGAGTTCCTCTGCTGCTGCG 259  
Db 13762 -----AAGCCGTAGAACGCCACACGGGCGAGTCTGAATGATGAATGGCTGATGGCGC 13813  
Qy 260 CCGCGAACTCTATCTAGCCCGCGCTTCTGCTGCGCACCCCTGCGTGGTACAGCGCGC 319  
Db 13814 AGATCTGATGTCGCCAGTCCCGGTATTTGCACTGGCGTCCATCTTAAAGCGCTGCGC 13873  
Qy 320 CGCGAAAGGCTGCGCATCTCCGGTGACATCATCTCTTCGCCCGCGAGGCGAAGGCCCC 379  
Db 13874 TGATGCGGAATCGAAATCGTTGGCATATGAGCTGTTCTGTCGGAAGCACAGCACC 13933  
Qy 380 GATCGTCGCCATCACCGGTTCCACGCGAAGACACCGTACCACCTGGTGGGCGAAAT 439  
Db 13934 GATTTGGCGGATTTACCGGTTCTTAACGCAAAACACGCTACCAACGCTAGTGGGTGAAAT 13993  
Qy 440 GCGCGTGGCGCGGACAGCGTGTGCGCTGCGCGGCAACCTCGCACCCCGCGGCTCGA 499  
Db 13994 GCGCAAGCGCGGGGTTACGTTGGTGGTGGCAATATTGGCTGCTGCTGCTGAT 14053  
Qy 500 CCGTCTGCGCGACACATCGAGCTGTAGTGTGGAGCTGTGAGCTTTCAGCTTTCAGTGAAC 559  
Db 14054 GCTACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 14113  
Qy 560 CTGGATCGCTCAACCGCGAGGTGGCGACCGTGTGACCTGACCTGACGAGACCATATGGA 619  
Db 14114 CACCTCCAGCTTACAGCGGTAGCAGCGACCATTTCTGAACGTGACTGAAGATCATATGA 14173  
Qy 620 TCGCTA---CGACGCGCATGGCTGACTACACCTGGCCAAAGCACCGGATCTTCCGCGGTGC 676  
Db 14174 TCGCTATCCGTTTGGTTTACACAGTATCGTGCAGCAAACTGCGCATTTACGAAACGC 14233  
Qy 677 CCGCCAGTGTGTGATTCGCGCGATGCGCTGACCGACCGCTGATGCGCGATACCGT 736  
Db 14234 GAAAGTTTGGTGGTTAATGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 14293  
Qy 737 GCGCTGCTGGTGTTCGCGCTGAAACAGCGGAGCTTCAAGCTTTCCGCGCTGATCGAGGA 796  
Db 14294 ACGCTGCTGAGCTTTGGGTGCAACATGGGTGATATCA-----CCTGAATATCA 14344  
Qy 797 AGACGGCAGAAAGTGGCTGGCTTCCAGTTTCGCAAGCTGCTGGCGTTGGCGAACTGAA 856  
Db 14345 GCAGGGCAACCTTGGCTGGCGGTTAAGCGGAGAAAGTGTGAATGTGAAGAGATGAA 14404  
Qy 857 GATCCGTGGCGCCACAACTATTCCAAAGCGCTGCGCGCTGGCGCTGGCGCTGCGCTGCG 916  
Db 14405 ACTTTCGGGCGAGCATAACTACACCAATGCGTGGCGCGCTGCGCTGGCAGATGCTGC 14464  
Qy 917 CGGCTGCGCTTCAGCGCCATGCTCGCGCGCTGAAGCGGCTTTCCGCGCTGGCTCATCG 976  
Db 14465 AGGTTTACCGCTGCCAGCGCTGAAAGCGCTTAACCACTTCACTGCTGCTGCGCGATCG 14524

```

Qy 977 CTGCCAGTGGGTACGCGAGCGGAGGCGGTGAGCTACTACGAGGATTCCAAAGGCCACCAA 1036
Db 14525 CTTTGAAGTTGTGCTGAGCATACGCGGTACGTTGGATTACGATTTCGAAAGCGACCAA 14584
Qy 1037 CGTCGGCGCGCCCTGCGCGGATCGAGGGGCTGGGTGCCGACATCGACGGCAAGTGGT 1096
Db 14585 CGTCGGCAGTACGGAAGCGCGCTGAATGCGCTG-----CAGGTAGACGGCACACTGCA 14638
Qy 1097 GCTGCTCGCGCGGAGAGCGCAAGGGCGCGGATTTCCATGACCTGCGCGAGCGGTCGC 1156
Db 14639 TTTGTTGCTGGGTGGCGATGGTAAATCGCGGACTTTAGCCCACTGGCGGTTACCTGAA 14698
Qy 1157 GCGCTTCGCCGGGCGGTGCTACTGCTTGGCCGTGAGCGCGGCTGATTCGCCAGGCACT 1216
Db 14699 TGGCGATAACGTACGTCTGTATTGTTTCGGTCTGTGACGGCGCGAGCTGGCGGCTAC- 14757
Qy 1217 GGGCAACGCGGTACCGCTGTGTGCGGCTGCGAAGCTGGACGAGCAGTCCGGGAGGCCGC 1276
Db 14758 -----GCCCGGAAGTGGCAGAACAAACCGAAACTATGGAACAGGCGATGCGCTTGCTGGC 14812
Qy 1277 CGAGCTGGCCCGGAGGCGATGCGTGTGTTGTCGCGGCGCTGCGGAGCCTGGACAT 1336
Db 14813 TCCGCGTGTTCAGCGGCGGATATGTTCTGCTCTCCCGACGCTGTGCCAGCCTTGATCA 14872
Qy 1337 GTTCAAGAACTTCGAAGAACGCGAGCGCTGTTGCGCAAGCCGTAGAGAGCTAGCGTG 1396
Db 14873 GTTCAAGAACTTGAACACGAGGCAATGAGTTTGCCCGTCTGGCGAAGGAGTTAGGTTG 14932
Qy 1397 ATGCTGTGCGGTGTGGCGCCCTTCCCGTCCGCGCTGTTGAGCCGGCACGGCATC 1450
Db 14933 ATGCGTTTATCTCTCCCTCGCCCTGAAATGCCCGCGCTGCCAGGATTTCAGTATC 14986

```

Search completed: August 14, 2003, 03:24:00  
Job time : 5318 secs